

HUMBOLDT-UNIVERSITÄT ZU BERLIN
INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT



BERLINER HANDREICHUNGEN
ZUR BIBLIOTHEKS- UND
INFORMATIONSWISSENSCHAFT

HEFT 270

**STAND UND PERSPEKTIVE DER ZEITUNGSDIGITALISIERUNG
IM INTERNATIONALEN VERGLEICH**

VON
ALEXANDER RINDFLEISCH

**STAND UND PERSPEKTIVE DER ZEITUNGSDIGITALISIERUNG
IM INTERNATIONALEN VERGLEICH**

**VON
ALEXANDER RINDFLEISCH**

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Konrad Umlauf
Humboldt-Universität zu Berlin

Heft 270

Rindfleisch, Alexander

Stand und Perspektive der Zeitungsdigitalisierung im internationalen Vergleich / von Alexander Rindfleisch. – Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2010. – 73 S. – (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 270)

ISSN 14 38-76 62

Abstract:

Zeitungen sind wichtige Quellen insbesondere der historischen Forschung. Ihre Digitalisierung würde die Benutzbarkeit erleichtern und verbesserte Recherchebedingungen schafft. Während in vielen Staaten große Projekte zur Zeitungsdigitalisierung existieren, stellen sie für Deutschland ein Desiderat dar. Daher wird hier – neben einem Überblick über die in Deutschland bereits existierenden Projekte – der Stand der Zeitungsdigitalisierung in den Niederlanden, Großbritannien und den USA untersucht. Diese Staaten weisen in einzelnen Bereichen vergleichbare Strukturen zu Deutschland auf und verfügen andererseits über umfangreiche, zentral koordinierte Zeitungsdigitalisierungsprogramme. Deren jeweilige Rahmenbedingungen und Strukturen werden hier untersucht, das heißt u. a. die Rolle einzelner Institutionen, die Finanzierung sowie die Digitalisierungsstrategien. Aufbauend auf diesem Ergebnis wird ein dreistufiges Konzept für ein nationales Zeitungsdigitalisierungsprogramm in Deutschland entworfen: In der ersten Phase der Digitalisierung soll aus zentralem Bestand ein Kern-Korpus digitalisierter historischer Zeitungen entstehen, in der zweiten Phase erfolgte eine konzertierte Digitalisierung in den Bundesländern und die dritte Phase bestünde in der Realisierung einer langfristig angelegten DFG-Aktionslinie „Zeitungsdigitalisierung“.

Diese Veröffentlichung geht zurück auf eine Master-Arbeit im postgradualen Fernstudiengang Master of Arts (Library and Information Science) an der Humboldt-Universität zu Berlin.

Online-Version:

<http://www.edoc.hu-berlin.de/series/berliner-handreichungen/2010-270>

Inhalt

A. Einleitung.....	7
B. Der Stand der Zeitungsdigitalisierung	11
1. Rahmenbedingungen und Projekte.....	11
a. Deutschland	11
b. Niederlande.....	19
c. Großbritannien.....	22
d. USA	27
e. Kommerzielle Zeitungsdigitalisierungsprojekte.....	32
2. Stand der Technik.....	35
a. Scan.....	35
b. Erschließung	39
c. Präsentation und Archivierung	44
3. Zusammenfassung und Einordnung	45
C. Die Perspektive: Ein Zeitungsdigitalisierungsprogramm für Deutschland	47
1. Grundlagen und Vorbereitungen	47
2. Erste Phase: Zentrale Digitalisierung	48
3. Zweite Phase: Dezentrale Digitalisierung	54
4. Dritte Phase: DFG-Aktionslinie „Zeitungsdigitalisierung“	56
D. Fazit	57
E. Literatur.....	59
1. Monographien, Zeitschriftenartikel, Online-Dokumente	59
2. Internetseiten & Blogs.....	70
3. Patente	73

A. Einleitung

„Historical newspapers are perhaps the single, most comprehensive resource on which to base a study of [... the past]. Newspapers have chronicled the daily life [...] in small towns and cities, since the first newspaper appeared.“¹

Politik, Alltag, Wirtschaft und Geistesgeschichte – die Lebenswirklichkeit einer bestimmten Zeit und Gesellschaft spiegelt sich in Zeitungen wider.² Diese waren lange Zeit die führenden Informationsträger und als solche meinungsbildend und für die Identität von Regionen und Gruppen prägend. Sie stellen daher eine zentrale und nicht durch andere Medien zu kompensierende Quelle für viele historische Fragestellungen dar.³ Das wird bereits an wenigen Beispielen deutlich: Politikhistoriker finden in lokalen Zeitungen Reden und Hinweise auf Auftritte von (landesweit bedeutenden) Politikern und können untersuchen, wie bestimmte Ereignisse in verschiedenen Regionen, in verschiedenen politischen oder weltanschaulichen Lagern gesehen wurden. Wirtschaftshistoriker werten Werbung, Geschäftsannoncen oder veröffentlichte Preislisten aus. Für die Alltags- und Kulturgeschichte bieten die ausschließlich in Zeitungen überlieferten Veranstaltungsanzeigen reiches Material. So sind beispielsweise Varieté- und Kinoprogramme für die film- oder theaterwissenschaftliche Forschung essentiell. Kulturhistorisch arbeitende Forschungsprojekte wie die Siegener Wanderkinodatenbank wären ohne möglichst lückenlos überlieferte Zeitungen nicht möglich. Der sprachwissenschaftlichen Forschung schließlich bietet sich durch Zeitungen die Möglichkeit fundierter synchroner oder diachroner Vergleiche zum Sprachgebrauch und zur Sprachentwicklung.⁴

Diesem Wert von Zeitungen als unentbehrlichen Quellen stehen jedoch größere Schwierigkeiten in der Benutzung gegenüber. Zeitungen sind in Bibliotheken zumeist als (Papier-) Original oder als Mikrofilm vorhanden. Sofern Mikrofilme vorliegen, werden die Originale den Nutzern in der Regel nicht bereitgestellt. Mikrofilm bringt für die Nutzer wiederum

¹ Bruce Cole: The National Digital Newspaper Program, OAH Newsletter, Mai 2004, <http://www.oah.org/pubs/nl/2004may/cole.html>. Für alle im Folgenden zitierten Onlinequellen gilt: letzter Zugriff 20.5.2009.

² Auch für diese Arbeit sollen die im Allgemeinen für die Charakterisierung von Zeitungen genannten Merkmale der Universalität, Periodizität, Aktualität und Publizität gelten. Vgl. Werner Faulstich: Zeitung, in: Werner Faulstich (Hrsg.): Grundwissen Medien, München 1994, S. 362-376. In dieser Arbeit wird primär die Digitalisierung „historischer“ Zeitungen behandelt. Darunter werden solche Zeitungen verstanden, die vor allem aufgrund ihres Erscheinungsdatums als gemeinfrei gelten.

³ Einen Überblick bietet: Alison Jones: The Many Uses of Newspapers, <http://dxs.richmond.edu/d/dds/docs/papers/usesofnewspapers.pdf>.

⁴ Ein gutes Praxisbeispiel für die Nutzung digitalisierter Zeitungen für die sprachwissenschaftliche Forschung bietet Barry Popik: Digital Historical Newspapers. A Review of the Powerful New Research Tools, in: Journal of English Linguistics 32 (2004) 2, S. 114-123.

einige Schwierigkeiten mit sich. Die Bedienung von Lesegeräten ist unter Umständen kompliziert, es kann nur eine sequentielle, relativ langsame „Suche“ erfolgen, die für Mikrofilme typische weiße Schrift auf schwarzem Grund ist für viele Nutzer schwer lesbar und die Anfertigung von Ausdrucken ist vergleichsweise kostspielig. Auch die Benutzung von Originalen ist mit Hindernissen verbunden: Bei der Anfertigung von Kopien entstehen auch hier hohe Kosten und hier wie dort sind Nutzer an die Räumlichkeiten und Öffnungszeiten der besitzenden Einrichtung gebunden.

Die Digitalisierung der Zeitungen kann in diesen Punkten Abhilfe schaffen. Der Zugang zu Zeitungen über das Internet entbindet die Nutzer von den Öffnungszeiten und Räumlichkeiten der Bibliotheken. Außerdem können digitalisierte Zeitungssseiten von Nutzern gespeichert und kostengünstig ausgedruckt werden.⁵ Die Digitalisierung schafft zugleich neue Möglichkeiten, die das Potential der Zeitungen als Quelle besser auszu-schöpfen vermögen. Dies gilt für Zeitungen in noch viel stärkerem Maße als für andere Drucke, beispielsweise Monographien oder Zeitschriften. Da Zeitungen nicht mono-thematisch sind, kein Inhaltsverzeichnis und auch keine klare, leicht zu erkennende Strukturierung aufweisen, eröffnen die Suche im Volltext und die Erschließung einzelner Artikel völlig neue Recherchemöglichkeiten. Hinzu kommt, dass Nutzer von Zeitungen zumeist nicht nur eine einzige Ausgabe durchsuchen wollen, sondern oftmals eine große Anzahl von aufeinander folgenden Ausgaben einer Zeitung benötigen oder verschiedene Zeitungen sichten. Hier schafft die Digitalisierung verbesserte Recherchebedingungen, da sie das gleichzeitige Durchsuchen mehrerer Zeitungstitel ermöglicht. Doch nicht nur für die Forschung beinhaltet die Digitalisierung von Zeitungen einen großen Mehrwert, sie kann auch der Bestandserhaltung dienen. Zeitungsdigitalisate ermöglichen es, die fragilen Originale der Benutzung zu entziehen und sie dadurch zu schonen. Da Zeitungen unbestreitbar einen Teil des kulturellen Erbes eines Landes ausmachen, trägt ihre Digitalisierung dazu bei, dieses Erbe unbeschadet zu überliefern.

Trotz dieses eindeutigen, großen Nutzens wurden in Deutschland bisher kaum Projekte zur Digitalisierung von Zeitungen realisiert, man konzentrierte sich zunächst vor allem auf Monographien und Zeitschriften. Hans Bohrmann nannte die Stellung von Zeitungen in Bibliotheken „ein Trauerspiel“, da es „heute noch erheblicher Anstrengungen [bedürfe], um den Zeitungen in Bibliotheken jenen Platz einzuräumen, der ihnen vom Benutzerinteresse her zukommt“⁶. Diese Aussage kann, mutatis mutandis, auch für den Stand der Zeitungsdigitalisierung in Deutschland gelten. Angesichts der erläuterten

⁵ Für die Bibliotheken hat die Digitalisierung darüber hinaus den Vorteil, dass hierdurch die Originale (oder die Mikrofilme) geschont und leichter konserviert werden können.

⁶ Hans Bohrmann: Zeitungen in deutschen Bibliotheken, in: Staatsbibliothek zu Berlin – Preußischer Kulturbesitz (Hrsg.): Zeitschriften in deutschen Bibliotheken: Bestand – Erwerbung – Erschließung, München u. a. 2005, S. 327-349, hier S. 327.

Bedeutung von Zeitungen für die Forschung und für die kulturelle Überlieferung besteht daher die dringende Notwendigkeit für ein landesweites strukturiertes Programm zur Zeitungsdigitalisierung in Deutschland.

International lässt sich dagegen bereits eine Vielzahl solcher groß angelegten Projekte zur Zeitungsdigitalisierung finden. Zeitungsdigitalisierungsprojekte werden im angloamerikanischen Raum ebenso durchgeführt wie in westeuropäischen Ländern (etwa in Frankreich, den Niederlanden oder Skandinavien). In diesen Ländern werden seit mehreren Jahren Erfahrungen mit der Digitalisierung umfangreicher Zeitungsbestände gesammelt, wobei inzwischen nicht nur kleinere Projekte durchgeführt, sondern Zeitungsdigitalisierungen in großem Umfang in landesweiten Programmen realisiert werden.

In der vorliegenden Arbeit werden diese in anderen Ländern gewonnenen Erfahrungen strukturiert aufbereitet, um sie für ein deutsches Zeitungsdigitalisierungsprogramm nutzbar zu machen. Die Untersuchung wird dabei in zwei Schritten erfolgen. In einem ersten Schritt wird untersucht, welche Rahmenbedingungen für die Digitalisierung von Zeitungen in Deutschland im Vergleich zu anderen, exemplarisch ausgewählten Ländern herrschen und welche Projektstrukturen innerhalb der jeweiligen Rahmenbedingungen entwickelt wurden. Der Vergleich wird sich neben Deutschland auf drei Staaten konzentrieren, die als beispielhaft für ein bestimmtes Vorgehen gelten können und zugleich Parallelen zu Deutschland aufweisen, die sie für einen Vergleich besonders interessant machen. Konkret werden Zeitungsdigitalisierungsprojekte in den Niederlanden, in Großbritannien und in den USA betrachtet. In den Niederlanden findet sich eine ähnliche Ausgangssituation wie in Deutschland, da in beiden Staaten eine Reihe von Institutionen damit begonnen hat, Teile ihrer Zeitungsbestände zu digitalisieren, ohne dass dem ein übergeordnetes Konzept zugrunde lag beziehungsweise liegt. Das Vereinigte Königreich hingegen kann als Musterbeispiel für eine zentral koordinierte Digitalisierung dienen. Die USA schließlich sind insofern ein interessantes Vergleichsland, da dort ebenso wie in Deutschland eine stark ausgeprägte föderale Struktur vorhanden ist. Der Vergleich von Zeitungsdigitalisierungsprogrammen in den genannten Ländern wird zeigen, wie in einzelnen Staaten bestimmte Fragen der Organisation einer umfassenden Zeitungsdigitalisierung gelöst wurden.

In einem zweiten Schritt wird anschließend der Stand der Technik des Digitalisierungs-Workflows betrachtet. Einzelne Aspekte der Arbeitsschritte Scan, Erschließung sowie Präsentation und Archivierung beeinflussen den Verlauf und die Kosten von Digitalisierungsprojekten maßgeblich. Daher sollen hier die Teilaspekte betrachtet werden, bei denen sich entweder Standards herausgebildet haben oder differierende, jedoch feststehende Alternativen vorhanden sind. Da sich technische Standards international herausbilden, sollen bei dieser Untersuchung die wesentlichen Entwicklungen weltweit berücksichtigt werden.

Auf diesen Ergebnissen aufbauend wird schließlich ein Konzept für ein Zeitungsdigitalisierungsprogramm in Deutschland entwickelt. Hierzu werden die in den beiden Untersuchungsschritten gewonnenen Erkenntnisse dahingehend betrachtet, wie ein solches Programm angesichts der konkreten Rahmenbedingungen in Deutschland konzipiert sein sollte. Die wesentlichen organisatorischen und technischen Fragen, die im Rahmen der Planung eines solchen Programms zu entscheiden sind, werden diskutiert, Lösungsvorschläge erarbeitet und es wird ein Grundlagenkonzept für ein Zeitungsdigitalisierungsprogramm in Deutschland vorgelegt.

Die Literaturlage zum Thema Zeitungsdigitalisierung stellt sich als recht spärlich dar. Weltweit existieren nur wenige Aufsätze – und keine Monographien – speziell zum Thema der Zeitungsdigitalisierung. Sicherlich ist Zeitungsdigitalisierung nur ein Teilaspekt des Themas Digitalisierung von Bibliotheks- und Archivgut. Da es jedoch in Standardwerken zur Digitalisierung wie „Digitizing Collections“ von Lorna Hughes⁷ nur marginal behandelt wird, wäre hier ergänzende Literatur wünschenswert, die auf die spezifischen Bedingungen der Zeitungsdigitalisierung eingeht. Vorhandene Literatur erschöpft sich meist in der Beschreibung einzelner Zeitungsdigitalisierungsprojekte, eine darüber hinausreichende oder systematische Analyse bildet ein Desiderat. Dies gilt auch für die Tagungsbände der „IFLA Newspaper Section“, in denen die Digitalisierung in den letzten Jahren zwar vermehrt Berücksichtigung findet, sich allerdings weitgehend in Projektdarstellungen erschöpft. Dem Gegenstand entsprechend finden sich die meisten Informationen online, einerseits in Form von E-Journals, vor allem aber auf den Projektseiten einzelner Digitalisierungsprojekte sowie zunehmend in Blogs. Die Literaturlage zur Zeitungsdigitalisierung speziell in Deutschland ist symptomatisch für die geringe Bedeutung, die diesem Thema hierzulande zukommt. Neben wenigen Aufsätzen zu einzelnen Projekten finden sich Informationen lediglich auf den Internetseiten einiger digitalisierender Institutionen.

⁷ Lorna Hughes: Digitizing Collections. Strategic Issues for the Information Manager, London 2004 (Digital Future Series).

B. Der Stand der Zeitungsdigitalisierung

1. Rahmenbedingungen und Projekte

Im Folgenden soll der Stand der Zeitungsdigitalisierung in Deutschland und den ausgewählten Vergleichsstaaten hinsichtlich der jeweiligen Ausgangslage und der derzeit zu konstatierenden Entwicklungen untersucht werden. Dabei werden jeweils zwei thematische Schwerpunkte im Vordergrund stehen: Erstens werden die in den betreffenden Staaten für die Digitalisierung von Zeitungen herrschenden Rahmenbedingungen betrachtet, wobei insbesondere folgende Fragen beantwortet werden: Wie viele Zeitungstitel sind für das jeweilige Land bekannt, in welchen Institutionen sind sie vorhanden, wie sind sie nachgewiesen und erschlossen? Welche bibliothekarischen oder archivarischen Strukturen in Bezug auf die Digitalisierung sind bereits vorhanden – im Allgemeinen und hinsichtlich der Zeitungen im Besonderen? Und schließlich: Wie sehen Finanzierungs- und Fördermöglichkeiten aus?

Zweitens werden bereits laufende und geplante Zeitungsdigitalisierungsprojekte betrachtet: Wie viele und welche Projekte existieren, welches Korpus, welche Zielsetzung und welches Vorgehen liegt ihnen zugrunde, wie werden die digitalisierten Zeitungen im Rahmen dieser Projekte erschlossen und zugänglich gemacht? Dabei sollen nicht nur einige herausragende oder größere Projekte für sich allein genommen vorgestellt, sondern gleichsam die Struktur innerhalb des gesamten Landes vor dem Hintergrund der jeweiligen Digitalisierungsstrategien bzw. „-policies“ dargelegt werden.

Anhand der genannten Leitfragen werden die wesentlichen Entwicklungen auf dem Sektor der Zeitungsdigitalisierung skizziert und im internationalen Vergleich die Merkmale des Vorgehens einzelner Staaten herausgearbeitet. Die Ergebnisse werden in einem späteren Schritt die Grundlage dafür bilden, die Perspektive der Zeitungsdigitalisierung in Deutschland zu evaluieren und die Leitlinien für ein Konzept einer landesweiten Zeitungsdigitalisierung in Deutschland zu entwerfen.

a. Deutschland

Deutschland besitzt, im Vergleich zu anderen westlichen Staaten, eine überaus reiche Zeitungslandschaft und Zeitungsgeschichte. Die weltweite erste Zeitung ist – für das Jahr 1604 – auf dem Gebiet des Reiches nachweisbar. Ende des 18. Jahrhunderts erschienen rund 200 Titel, Ende des 19. Jahrhunderts waren es bereits über 1.900 und in der Weimarer Republik über 4.200. Eine genaue Übersicht über die derzeit in deutschen Bibliotheken, Archiven, Museen und anderen Einrichtungen vorhandenen Bestände liegt jedoch (immer

noch) nicht vor. Zwar existiert eine Reihe von bibliographischen Übersichten, diese sind jedoch entweder zeitlich oder thematisch eng gefasst oder durch die Bestands- und territorialen Verschiebungen infolge des Zweiten Weltkriegs überholt.⁸ Und auch wenn sich inzwischen ein viel versprechender Ansatz einer landesweiten Titel- und Bestandsübersicht findet, da die Zeitschriftendatenbank (ZDB) mittlerweile auch rund 90.000 Zeitungstitel nachweist – so erfolgt jedoch auch hier keine systematische Bestandsaufnahme.⁹ Darüber hinaus wird an der Staatsbibliothek zu Berlin der Aufbau einer deutschen Zeitungstopographie betrieben, in der die in den verschiedenen Gedächtnisinstitutionen und anderen Einrichtungen vorhandenen Bestände zumindest quantitativ erfasst werden sollen.¹⁰ Mit Hilfe dieser beiden Instrumente und unter Hinzuziehung älterer Bestandsverzeichnisse ist es möglich, zumindest in Ansätzen einen Überblick über die Zeitungsbestände in Deutschland zu erhalten. Trotz des Nutzens dieser Nachweisinstrumente bilden sie jedoch keinen Ersatz für ein – vielfach gefordertes, bisher aber nicht realisiertes – nationales Erschließungsprogramm. Eine systematische Verzeichnung, wie sie in anderen Staaten durchgeführt wurde, bleibt daher ein Desiderat.

Der Grund für die problematische Nachweissituation ist zum einen in den Weltkriegen und den dadurch bedingten Zerstörungen, Bestandsverlagerungen und territorialen Verschiebungen zu sehen. Zum anderen ist sie eine Folge der historisch bedingten „Zersplitterung“ Deutschlands, der Vielzahl souveräner Regionen und der großen Bedeutung der einzelnen Regionen. Diese Aufteilung führte zu einer an Zahl und Formen reichen Zeitungslandschaft, bedingte aber zugleich deren problematische Überlieferung. Eine umfassende systematische Sammlung von Zeitungen durch Bibliotheken fand in Deutschland bis 1912 ebenso wenig statt wie die monographischer Werke. Daran änderte auch die Einrichtung der Deutschen Bücherei in Leipzig 1912 nichts, da diese von Anfang an nicht den Auftrag besaß, Zeitungen zu sammeln. Die Deutsche Bibliothek schließlich, beziehungsweise nunmehr die Deutschen Nationalbibliothek, sammelt diese Zeitungen nur selektiv und lediglich auf Mikrofilm. Man geht davon aus, die regionalen Zeitungen würden (vor allem im Rahmen der Pflichtabgabe) in den Landesbibliotheken gesammelt.

⁸ Hier sind zu nennen: Hans Traub: Standortskatalog wichtiger Zeitungsbestände in deutschen Bibliotheken, Leipzig 1933; Gerd Hagelweide: Deutsche Zeitungsbestände in Bibliotheken und Archiven, Düsseldorf 1974. Vgl. insgesamt: Hartmut Walravens: Newspaper Cataloguing in Germany, in: Hartmut Walravens/Edmund King (Hrsg.): Newspapers in International Librarianship. Papers presented by the Newspapers Section at IFLA Conferences (IFLA Publications, Bd. 107), S. 71-75; sowie die Beiträge in dem Sammelband: Hans Bohrmann/Wilbert Ubben (Hrsg.): Zeitungen verzeichnen und nutzen. Aktuelle Ansätze und Unternehmungen zur bibliographischen und archivalischen Beschreibung und Nutzung deutschsprachiger Zeitungen, Berlin 1998 (Informationsmittel für Bibliotheken, Beiheft 7).

⁹ Stand: 2008. Vgl. Hartmut Walravens: Bibliographic Control of Newspaper Collections, in: Hartmut Walravens (Hrsg.): Newspaper Collection Management: Printed and Digital Challenges. Proceedings of the International Newspaper Conference, Santiago de Chile, April 3-5, 2007, München 2008 (IFLA Publications, Bd. 133), S. 265-273.

¹⁰ Vgl. <http://daten.zeitschriftendatenbank.de/projekte/zeitungstopographie/allgemein.html>.

Doch auch dies ist nicht durchgängig der Fall.¹¹ Lediglich für die ausländische Presse wurde das Sondersammelgebiet „Ausländische Zeitungen“ an der Staatsbibliothek zu Berlin eingerichtet.¹²

Somit ergibt sich hinsichtlich der Digitalisierung von Zeitungen folgende strukturelle Ausgangssituation in Deutschland: In der Nationalbibliothek finden sich lediglich Teile der Zeitungsbestände mit Erscheinungsjahr ab 1913, vor allem aber ab 1969, in Auswahl und als Mikrofilm. Große – auch historische – Zeitungsbestände sind vor allem in der Staatsbibliothek zu Berlin und in anderen Staats- sowie Landesbibliotheken vorhanden, unter anderem in München, Bremen oder Münster.¹³ Daneben finden sich umfangreichere Zeitungsbestände im „Mikrofilmarchiv der Deutschsprachigen Presse“ (MFA) in Dortmund. Institutionen, die in einem der breit angelegten DFG-Programme zur Mikroverfilmung von Zeitungen und anderem gefährdeten Bibliotheksgut gefördert wurden, sind verpflichtet, von den im Rahmen dieser Programme angefertigten Filmen eine Master-Kopie an das MFA abzugeben.¹⁴ Schließlich weisen viele (gerade auch kleinere) Regionalbibliotheken, vor allem aber auch Archive und Museen einen umfangreichen Bestand an andernorts kaum oder gar nicht vorhandenen Titeln auf.¹⁵

Nach dieser Übersicht über das Quellmaterial für eine eventuelle Zeitungsdigitalisierung soll im Folgenden betrachtet werden, welche Strukturen hinsichtlich der Digitalisierung von Bibliotheks- und Archivgut in Deutschland grundsätzlich bestehen. Seit den Anfängen der Digitalisierung in den 1990er Jahren haben Gedächtniseinrichtungen in Deutschland viele Erfahrungen mit verschiedenen Techniken und Verfahren der Digitalisierung gesammelt. Doch trotz der gewonnenen Kenntnisse, der Weiterentwicklung der Technik und der insgesamt deutlich gesunkenen Kosten bei gesteigerter Qualität ist die Digitalisierung immer noch ein zeit- und geldintensiver Prozess. Viele Institutionen können und/oder wollen sich diesen Aufwand nicht leisten; die Zahl der tatsächlich in größerem Maße digitalisierenden Einrichtungen ist daher nach wie vor gering. Insbesondere der Aufbau einer eigenen technischen Infrastruktur, das heißt einer eigenen Digitalisierungsstelle, wird nur von wenigen Einrichtungen betrieben. Neben einigen größeren Universitäts- und Landesbibliotheken sind hier insbesondere die

¹¹ Vgl. Catharina Sodann/Susann Solberg: Zeitungsbestände in der Deutschen Nationalbibliothek, in: Zeitungs-Mikrofilm-Nachrichten 10 (2008) 11, S. 6-10.

¹² Vgl. <http://webis.sub.uni-hamburg.de/webis/index.php/26>.

¹³ Vgl. auch: Historische Zeitungsbestände und ihre Verfilmung: Studie zur Bestimmung des Anteils der noch nicht verfilmten Zeitungsbestände in Deutschland, erarb. von Manfred Pankratz, Berlin 1994 (Schriften der Deutschen Forschungsgemeinschaft).

¹⁴ Vgl. Reinhold Altenhöner: Neues Förderungsprogramm der Deutschen Forschungsgemeinschaft zur Mikroverfilmung gefährdeter Bibliotheksbestände, in: Zeitschrift für Bibliothekswesen und Bibliographie 40 (1993) 3, S. 254-260.

¹⁵ Vgl. Hans Bohrmann: Zeitungen in Bibliotheken – ein Trauerspiel? In: Bibliothek. Forschung und Praxis 54 (2003) 3, S. 162-164, hier S. 164. Gerade die in den Archiven und Museen vorhandenen Bestände sind kaum verzeichnet.

Niedersächsische Staats- und Universitätsbibliothek Göttingen und die Bayerische Staatsbibliothek München zu nennen. Hier wurden 1997 Digitalisierungszentren eingerichtet, die als „nationale Kompetenzzentren“ eine mehrfache Aufgabe wahrnehmen: als Digitalisierungsdienstleister, als beratende Serviceeinrichtungen sowie als Forschungs- und Entwicklungszentren.¹⁶ Darüber hinaus werden inzwischen auch in anderen Bundesländern beziehungsweise Bibliotheksverbünden Digitalisierungskompetenzen gebündelt und an zentraler Stelle vorgehalten. So bietet das Hochschulbibliothekszentrum (hbz) in Köln für die Bibliotheken des nordrhein-westfälischen Bibliotheksverbundes Dienstleistungen und Infrastrukturen an,¹⁷ ebenso versteht sich das Landesbibliothekszentrum Rheinland-Pfalz (LBZ) als Serviceeinrichtung mit der Aufgabe, auch im Bereich der Digitalisierung rheinland-pfälzische Bibliotheken zu unterstützen.

Daneben existiert eine Reihe kommerzieller Anbieter, die sich auf die Digitalisierung von Bibliotheks- und Archivgut spezialisiert haben. Doch anders, als dies beispielsweise in den USA der Fall ist, verfolgen diese Firmen keine selbständigen Digitalisierungsprojekte, sondern bieten Dienstleistungen an, vom einfachen Scannen bis hin zur kompletten Projektabwicklung.¹⁸

Während somit Bestand und Digitalisierungskompetenz in mehr oder weniger großem Maße dezentral vorhanden sind, liegt die Finanzierung – insbesondere großer – Digitalisierungsvorhaben primär in der Verantwortung weniger Institutionen. Digitalisierungsprojekte sind teuer und aus den regulären Etats der besitzenden Einrichtungen vielfach nicht zu finanzieren. Voraussetzung für die Realisierung solcher Maßnahmen ist daher zumeist die Bewilligung von Sonder- beziehungsweise Projektmitteln. Neben Mitteln aus Förderprogrammen der Bundesländer oder anderer öffentlicher Einrichtungen (u. a. Stiftungen) sowie Zuwendungen des Bundesministeriums für Bildung und Forschung (BMBF) kommt hierfür vor allem eine Förderung durch die Deutsche Forschungsgemeinschaft (DFG) in Frage.

Die DFG hat die Grundlagen ihrer Digitalisierungspolitik in verschiedenen Positionspapieren dargelegt. Im DFG-Positionspapier „Ziele und Struktur des Förderprogramms Kulturelle Überlieferung“ von 2005 heißt es: „Die gemeinfreie nationale kulturelle Überlieferung sollte in ihrer Gesamtheit erschlossen, dokumentiert und digital zur Verfügung gestellt werden.“¹⁹ Und weiter: Das Förderprogramm „Kulturelle Überlieferung

¹⁶ Vgl. <http://www.digitale-sammlungen.de/index.html?c=mdz10&l=de>.

¹⁷ Vgl. <http://www.hbz-nrw.de/angebote/hosting/scantoweb>.

¹⁸ Eine Übersicht über entsprechende Dienstleister findet sich beispielsweise im Forum Bestandserhaltung: <http://www.forum-bestandserhaltung.de>.

¹⁹ Deutsche Forschungsgemeinschaft: DFG-Positionspapier: Ziele und Struktur des Förderprogramms Kulturelle Überlieferung, Bonn 2005, S. 2.

hilft beim Aufbau einer frei zugänglichen nationalen digitalen Forschungsbibliothek“.²⁰ Die vor diesem Hintergrund bisher geförderten Projekte weisen jedoch einen begrenzten thematischen Fokus auf. Landesweite systematische Projekte sind – mit der gleich zu nennenden Ausnahme – im Rahmen des Förderschwerpunkts „Retrospektive Digitalisierung von Bibliotheksbeständen“ nicht zu finden.²¹ Zugleich besteht eine Diskrepanz zwischen dem programmatisch formulierten Ziel und der Förderungspraxis, da für die Förderung eines Projekts zumeist der Nachweis eines direkten Nutzens für die Forschung notwendig ist. Eine Förderung der Digitalisierung des „kulturellen Erbes“ ohne direkte Forschungsanbindung ist schwieriger zu erreichen. Symptomatisch heißt es in der aktuellen Fassung des Merkblattes „Erschließung und Digitalisierung handschriftlicher und gedruckter Überlieferung“ auch: „Ziel ist die Erschließung und/oder Digitalisierung herausragender und für die Forschung überregional bedeutender Bestände der handschriftlichen und/oder gedruckten Überlieferung.“²²

Eine Ausnahme stellt das Verzeichnis der deutschsprachigen Drucke des 18. Jahrhunderts (VD18) dar. Im Unterschied zu VD16 und VD17 sollen im Rahmen des VD18 die erschlossenen Drucke gleichzeitig digitalisiert werden.²³ Für Zeitungen hingegen wurde bis dato von der DFG kein entsprechendes Programm aufgelegt. Dies wird auch durch die Tatsache illustriert, dass die DFG-Praxisregeln für die Durchführung von Digitalisierungsprojekten vor allem auf die Digitalisierung von Monographien, anderen Einzelwerken oder Zeitschriften abheben, die Charakteristika und Probleme der Zeitungsdigitalisierung jedoch keine Berücksichtigung finden. Abschließend lässt sich daher feststellen, dass in Deutschland zum gegenwärtigen Zeitpunkt keine national koordinierte Digitalisierung von Zeitungen existiert.

Die einzigen existierenden Maßnahmen sind eine Reihe von regional oder thematisch begrenzten Projekten. Diese sollen im Folgenden vorgestellt werden, da sie zusammen genommen den Stand der Zeitungsdigitalisierung in Deutschland wiedergeben:

Im Rahmen der Bayerischen Landesbibliothek Online (BLO) wurden bisher acht Zeitungen digitalisiert und online präsentiert, zwei weitere sind in Vorbereitung. Verantwortlich für die Digitalisierungen waren jeweils die besitzenden bayerischen Bibliotheken in Kooperation mit der Bayerischen Staatsbibliothek (BSB). Alle Zeitungen sind über eine zentrale Einstiegsseite der BLO erreichbar. Es werden keine

²⁰ http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/projektfoerderung/foerderziele/index.html.

²¹ Vgl. hierzu: Manfred Thaller u. a.: „Retrospektive Digitalisierung von Bibliotheksbeständen“ – Evaluierungsbericht über einen Förderschwerpunkt der DFG, Universität zu Köln, Januar 2005.

²² Deutsche Forschungsgemeinschaft: Merkblatt: Erschließung und Digitalisierung handschriftlicher und gedruckter Überlieferung, Bonn 2009.

²³ Diese Drucke umfassen neben Monographien und Zeitschriften potentiell auch Zeitungen, diese werden jedoch wohl nur eine marginale Bedeutung haben.

Suchmöglichkeiten geboten, eine Recherche ist über das Datum beziehungsweise die Zeitungsausgabe möglich.²⁴

Darüber hinaus wurde in Bayern in einem DFG-geförderten Projekt eine Bestandsaufnahme und Verzeichnung der in bayerischen Bibliotheken, Archiven und anderen Einrichtungen vorhandenen Zeitungen vorgenommen, die zugleich als Ausgangspunkt für ein Digitalisierungsunternehmen dienen soll.²⁵ Ein detaillierter Plan liegt derzeit noch nicht vor, nach Angaben der BSB sieht das Konzept jedoch ein Vorgehen in drei Stufen vor. Zunächst ist die Digitalisierung von 25 Titeln (insgesamt rund 4 Millionen Seiten) geplant, die eine weite Verbreitung aufwiesen und vor allem in den regionalen Zentren (München, Augsburg u. a.) publiziert wurden. In der zweiten Stufe sollen im Rahmen lokaler Projekte regionale und lokale Zeitungen aus mehr als 30 Publikationsorten digitalisiert werden. Die dritte Stufe schließlich sieht die Digitalisierung bestimmter thematischer Gruppen von Zeitungen vor. Aufgrund der Problematik, Fraktur-Schriften mittels OCR-Software zu erfassen,²⁶ ist eine Erschließung lediglich auf Ebene der Ausgaben vorgesehen – letztlich also auf der gleichen Ebene, wie sie sich bereits in der BLO findet. Letztere stellt auch für das geplante Programm die in Aussicht genommene Präsentationsoberfläche dar. Für die Finanzierung der Projektphase liegen noch keine konkreten Pläne vor. Beabsichtigt ist neben der Beantragung von DFG-Mitteln eine fallweise Finanzierung auf lokaler Ebene.

Die Staatsbibliothek zu Berlin digitalisiert in Kooperation mit dem Institut für Zeitungsforschung in Dortmund sowie dem Institut für Kommunikationswissenschaft der Universität Bamberg die Amtspresse Preußens. Das erklärte Ziel des Projekts ist es, „die auflagenstärkste und einflussreichste politische Presse des letzten Drittels im 19. Jahrhundert“ zu digitalisieren, mit einem Kommentar zu versehen und zu indizieren. In zwei ersten Projektphasen wurden die Provinzial-Correspondenz (1863-1884) und ihr Quasi-Nachfolger, die Neuesten Mittheilungen (1882-1894), bearbeitet.²⁷ Derzeit befindet sich der dritte und letzte Titel, das Teltower Kreisblatt, in der Digitalisierungsphase. Hierbei geht die Erschließung über die der beiden Vorgängerprojekte hinaus; in Kooperation mit dem Bureau Ingénieur Tomasi wird dessen OCR-Software BIT-Fraktur zur Volltexterkennung erprobt. Darüber hinaus plant die Staatsbibliothek zu Berlin derzeit

²⁴ Vgl. <http://www.bayerische-landesbibliothek-online.de/zeitungen/>.

²⁵ Vgl. Richard Mai/Hildegard Schäffler: Bayerische Zeitungen und Amtsblätter bis zum Ende des Zweiten Weltkrieges – Abschluss eines DFG-geförderten Erschließungsprojekts und Entwurf eines Digitalisierungskonzepts, in: Zeitungs-Mikrofilm-Nachrichten 10 (2008) 11, S. 10-13; Klaus Ceynowa: Digitising Historic Newspapers in Germany – the Case of Bavaria, in: Hartmut Walravens (Hrsg.): The Impact of Digital Technology on Contemporary and Historic Newspapers. Proceedings of the International Newspaper Conference, Singapore, 1.-3. April 2008, and Papers from the IFLA World Library and Information Congress, Québec, Canada, August 2008, München 2008 (IFLA Publications, Bd. 135), S. 39-46.

²⁶ Vgl. dazu Kapitel A.2.b.

²⁷ <http://amtspresse.staatsbibliothek-berlin.de/projekt.html>.

zusammen mit dem Zentrum für Zeithistorische Forschung Potsdam die Digitalisierung von Tageszeitungen der DDR. Hierzu sollen zunächst drei Titel digitalisiert und erschlossen sowie eine Online-Recherche- und Forschungsumgebung aufgebaut werden.²⁸

Das Landesbibliothekszenrum (LBZ) Rheinland-Pfalz und die UB Trier betreiben seit 2007 gemeinsam das Portal dilibri als „einheitliches Digitalisierungsportal für Rheinland-Pfalz“.²⁹ Primär ist dilibri als Sammlung digitalisierter landeskundlicher Werke zu Rheinland-Pfalz konzipiert. In diesem Rahmen wurden bisher (in Kooperation mit externen Dienstleistern) auch 30 Zeitungen digitalisiert – überwiegend Fastnachtszeitungen sowie humoristische, satirische Zeitungen.³⁰

Im Rahmen des Projekts „Exilpresse digital“ wurden von 1998 bis 2003 in der Deutschen Bibliothek ausgewählte Exilzeitungen und -zeitschriften aus dem Bestand des „Deutschen Exilarchivs 1933-1945 Frankfurt am Main“ sowie der „Sammlung Exil-Literatur Leipzig“ digitalisiert – insgesamt 30 Zeitungen mit rund 100.000 Seiten, „die ein möglichst breites Spektrum der Exilpresse 1933-1945 präsentieren sollen“.³¹ Neben dem Blättern durch die einzelnen Zeitungen ist die Suche nach Titeln, Verfassern sowie im gesamten Volltext möglich. Für die Anzeige kann zwischen Bilddateien in mehreren Auflösung sowie dem PDF-Format gewählt werden. Gefördert wurde das Projekt durch die DFG.

Das Portal „Compact Memory“ bietet umfangreiche digitalisierte Bestände jüdischer Periodika zwischen 1806 und 1938. Das Projekt wurde durch das Lehr- und Forschungsgebiet „Deutsch-Jüdische Literaturgeschichte“ der RWTH Aachen und die Universitätsbibliothek Frankfurt a. M. (mit dem Sondersammelgebiet Judentum) in Kooperation mit der Bibliothek Germania Judaica in Köln realisiert. Insgesamt wurden 118 Periodika, darunter jedoch nur wenige Zeitungen, digitalisiert.³²

²⁸ Vgl. [http://zefys.staatsbibliothek-berlin.de/index.php?id=newsartikel&tx_ttnews\[tt_news\]=13&tx_ttnews\[backPid\]=24&cHash=baaeec751f](http://zefys.staatsbibliothek-berlin.de/index.php?id=newsartikel&tx_ttnews[tt_news]=13&tx_ttnews[backPid]=24&cHash=baaeec751f).

²⁹ Vgl. Barbara Koelges u. a.: dilibri – ein einheitliches Digitalisierungsportal für Rheinland-Pfalz, in: BIT online 11 (2008) 3, S. 276-279.

³⁰ Die Fastnachtszeitungen können zwar, werden die in der Einleitung genannten Kriterien zugrunde gelegt, nicht als Zeitungen im eigentlichen Sinne verstanden werden, sie sollen aufgrund der Materialähnlichkeit hier aber gleichwohl genannt werden.

³¹ <http://deposit.ddb.de/online/exil/exil.htm>; vgl. auch Renate Seib: Exilpresse digital: The Deutsche Bibliothek's Digitization of Selected German Exile Periodicals and Newspapers from the 1933-1945 Period, in: The Serials Librarian 43 (2003) 2, S. 29-39.

³² Vgl. <http://www.compactmemory.de/>; Hans Otto Horch: Compact Memory – ein DFG-Projekt zur retrospektiven Digitalisierung jüdischer Periodika im deutschsprachigen Raum, in: Zeitschrift für Bibliothekswesen und Bibliographie, 53 (2006), S. 177-180; Rachel Heuberger: Bewahren und Erschließen: die deutschsprachige jüdische Presse und das Projekt Compact Memory, Vortrag auf der Konferenz .hist 2006, Berlin 2006, <http://publikationen.ub.uni-frankfurt.de/volltexte/2007/4857>.

Die UB Heidelberg hat 19 der im Bestand vorhandenen Feldzeitungen mikroverfilmt und digitalisiert. Lücken im eigenen Bestand wurden dabei durch Bestände anderer Institutionen ergänzt.³³

Abschließend sind zwei kleinere Projekte zu nennen: In der ULB Halle werden im Rahmen eines DFG-geförderten Projekts zwei regionale Zeitungen – das Naumburger Kreisblatt sowie die Wettinger Zeitung – digitalisiert. An der SLUB Dresden werden im Rahmen der Digitalisierung verschiedener Sammlungen auch einige regionale Zeitungen digitalisiert.³⁴

Über diese, durch öffentliche Einrichtungen durchgeführten Projekte hinaus digitalisieren einige Verlage retrospektiv ihre Archive, stellen die digitalen Dokumente aber nur in seltenen Fällen (kostenlos) öffentlich zur Verfügung.³⁵

Mit diesen Projekten ist der gegenwärtige Stand der Zeitungsdigitalisierung in Deutschland beschrieben. Dabei lässt sich feststellen, dass erste und wichtige Ansätze für die Digitalisierung von Zeitungen in Deutschland bereits unternommen und dabei Erfahrungen mit der Technik und den Verfahren gesammelt wurden. Ungeachtet des Wertes der so gewonnenen Kenntnisse und der durchgeführten Projekte ist der Stand der Zeitungsdigitalisierung in Deutschland – von einer übergeordneten nationalen Perspektive aus betrachtet – jedoch defizitär. Für eine Reihe von Vorhaben existieren erst allgemeine Konzepte, andere befinden sich noch in der Planungsphase. Die bereits abgeschlossenen Projekte umfassen mehr oder weniger große und auch wichtige Korpora. Sie sind jedoch regional (Bayern, Rheinland-Pfalz), zeitlich (Exilpresse) und/oder thematisch eng begrenzt (jüdische Presse, DDR-Zeitungen, preußische Amtspresse). Die Einordnung der bereits erfolgten Zeitungsdigitalisierung wird schon bei der Betrachtung einiger Zahlen deutlich. Die bereits abgeschlossenen Digitalisierungsprojekte umfassen rund 100 digitalisierte Zeitungen – eine verschwindend kleine Anzahl angesichts mehrerer 10.000 jemals publizierter deutscher Zeitungstitel.

Hinzu kommt schließlich, dass vielfach Digitalisierungsprojekte und die daraus resultierenden Bestände nicht oder nur schwierig zu finden sind. Zwar sind beispielsweise DFG-geförderte Projekte verpflichtet, ihre Daten über relevante Portale im Internet zugänglich zu machen. Es existiert auch eine Reihe von Portalen, insbesondere das Zentralverzeichnis Digitalisierter Drucke (zvdd) – sie werden jedoch nicht in ausreichendem Maße genutzt.³⁶ Gerade für digitalisierte Zeitungen wäre ein einheitlicher Sucheinstieg über ein zentrales Portal notwendig, das zum gegenwärtigen Zeitpunkt allerdings nicht existiert. Es bleibt abzuwarten, ob beispielsweise das von der

³³ Vgl. <http://feldzeitungen.uni-hd.de/>.

³⁴ Vgl. <http://bibliothek.uni-halle.de/dbib/digital/zeitungen/>; <http://digital.slub-dresden.de/sammlungen/>.

³⁵ Ein Beispiel hierfür ist das Archiv der ZEIT: <http://www.zeit.de/archiv/index>.

³⁶ Vgl. <http://www.zvdd.de/index.html>.

Staatsbibliothek zu Berlin installierte Zeitungsportals (ZEFYS) zukünftig als zentraler Sucheinstieg dienen kann.

b. Niederlande

Trotz der wesentlich geringeren Größe des Landes ähneln die Rahmenbedingungen für die Zeitungsdigitalisierung in den Niederlanden den deutschen in einer Reihe von Punkten. Die darauf aufbauenden Entwicklungen unterscheiden sich jedoch. Beides soll im Folgenden erläutert werden.

Insgesamt wird von rund 7.000 in den Niederlanden jemals publizierten Zeitungstiteln ausgegangen, eine umfassende, vollständige Verzeichnung dieser Zeitungen existiert jedoch, ebenso wie in Deutschland, nicht.³⁷ Die Königliche Bibliothek in Den Haag, die Nationalbibliothek der Niederlande, besitzt kein Pflichtexemplarrecht. Sie fungiert jedoch als nationale Depotbibliothek, so dass niederländische Monographien und Periodika, darunter auch Zeitungen, von den Verlagen freiwillig an sie abgegeben werden.³⁸ Darüber hinaus sind Zeitungen in vielen anderen Institutionen, zumeist Bibliotheken, Archiven und Museen, vorhanden. Entsprechend diesem verteilten vorhandenen Bestand werden auch in den Niederlanden an zahlreichen Bibliotheken und anderen Einrichtungen Zeitungen digitalisiert. Ein Großteil der Institutionen begann dabei wesentlich früher mit der Digitalisierung ihres Zeitungsbestandes als die Königliche Bibliothek, so dass sich insgesamt – ähnlich wie in Deutschland – eine äußerst heterogene Situation ergab. Im Unterschied zu Deutschland wird inzwischen in den Niederlanden der Versuch unternommen, dieser Heterogenität durch eine Konzentration und Vereinheitlichung der Projekte zu begegnen. Hierzu hat die Königliche Bibliothek in Den Haag zwei Vorhaben begonnen: Zum einen wurde ein Massendigitalisierungsprojekt für niederländische Zeitungen initiiert, zum anderen hat sich auf Initiative der Königlichen Bibliothek ein nationaler Expertenrat konstituiert, der auf eine Vereinheitlichung und Koordinierung technischer Standards und Verfahren hinarbeiten soll. Beide Ansätze sind für eine mögliche weitere Entwicklung der Zeitungsdigitalisierung in Deutschland aufgrund der ähnlichen Ausgangssituation besonders interessant und sollen daher im Folgenden betrachtet werden.

Die Königliche Bibliothek fungiert als Nationalbibliothek zugleich als niederländisches Kompetenzzentrum für Digitalisierung. In ihrer „Beleid digitalisering“ nennt sie drei

³⁷ Vgl. Astrid Verheusen: Newspaper Digitisation in the Netherlands. The Dutch Digital Databank for Newspapers and Other Initiatives, in: Hartmut Walravens (Hrsg.): The Impact of Digital Technology on Contemporary and Historic Newspapers. Proceedings of the International Newspaper Conference, Singapore, 1.-3. April 2008, and Papers from the IFLA World Library and Information Congress, Québec, Canada, August, 2008, München 2008 (IFLA Publications, Bd. 135), S. 55-61, hier S. 56; Astrid Verheusen: Databank Digitale Dagbladen, in: edata & research 1 (2007) 4, S. 4.

³⁸ Koninklijke Bibliotheek: Annual report 2007, Den Haag 2008, S. 16.

wesentliche Ziele der Digitalisierung: Zum einen sollen Dokumente leichter zugänglich gemacht werden, zum zweiten werden neue und effektivere Möglichkeiten der Recherche geschaffen und schließlich diene die Digitalisierung auch der Bestandserhaltung.³⁹ Ausgehend von diesen Aspekten und vor dem Hintergrund sich weiterentwickelnder Technik wurden zu Beginn der Digitalisierungsaktivitäten vor allem ausgewählte, für die Öffentlichkeit (visuell) besonders attraktive Einzelstücke digitalisiert, gefolgt von Querschnittsprojekten zum niederländischen Kulturerbe. Die erste Massendigitalisierung erfolgte ab 2003 mit der Digitalisierung der kompletten niederländischen Parlamentsschriften.

Im Rahmen des zweiten großen Massendigitalisierungsprojekts wurden erstmals in größerem Umfang Zeitungen digitalisiert und die „Databank Digitale Dagbladen“ konzipiert. Das Projekt wurde Ende 2006 begonnen, soll bis 2011 laufen und die Digitalisierung von rund acht Millionen Zeitungsseiten umfassen.⁴⁰ Die Finanzierung in Höhe von 12,5 Millionen Euro erfolgt über das Programm „Nationaal programma voor investeringen in grootschalige onderzoeksfaciliteiten“ der „Nederlandse Organisatie voor Wetenschappelijk Onderzoek“ (NWO), der niederländischen Einrichtung zur Wissenschaftsförderung.⁴¹

Das Korpus soll nationale, regionale, lokale und koloniale Zeitungen vom 17. bis zum 20. Jahrhundert umfassen und somit einen umfassenden Querschnitt der Zeitungspublikation und des Lebens in den Niederlanden bieten. Zur Auswahl der Zeitungstitel wurde ein mehrstufiges Verfahren gewählt. Zunächst wurden sechs Zeitabschnitte gebildet und für jeden die bestimmenden politischen, gesellschaftlichen, ökonomischen und kulturellen Charakteristika sowie die Entwicklung von Presse und Zeitungswesen erarbeitet. Aus diesen Merkmalen wurden für jede Periode Auswahlkriterien zusammengestellt, wobei nach Angaben der Projektverantwortlichen das Ziel verfolgt wurde, solche Zeitungen auszuwählen, die gesellschaftliche Entwicklungen besonders widerspiegeln oder selbst beeinflusst hätten.⁴² In einem zweiten Schritt wurden praktische Aspekte berücksichtigt, unter anderem der Aufwand, komplette Erscheinungsverläufe eines Titels zur Verfügung zu stellen, die Qualität der Mikrofilme sowie Urheberrechtsfragen. Anhand dieser Kriterien wurde – unter Beteiligung eines beratenden Gremiums aus (Presse-)Historikern – die endgültige Auswahl getroffen, die letztendlich 48 Titel umfasste.

³⁹ Vgl. <http://www.kb.nl/hrd/digitalisering/beleid-en.html>.

⁴⁰ Vgl. <http://www.kb.nl/hrd/digitalisering/beleid-en.html>; <http://www.kb.nl/hrd/digi/ddd/index-en.html>.

⁴¹ Vgl. Koninklijke Bibliotheek: Annual report 2007, S. 21; <http://www.nwo.nl/projecten/nsf/pages/2300133170>.

⁴² Hiervon abweichend wurden für die kolonialen Zeitungen modifizierte Kriterien erarbeitet. Für die Zeit des Zweiten Weltkriegs konnten weitere Fördermittel akquiriert werden, so dass hier zumindest alle damals illegal erschienenen Zeitungen komplett digitalisiert werden können.

In Vorbereitung auf das Projekt fanden von 1999 bis 2004 zwei Pilotstudien statt, deren Finanzierung im Rahmen des nationalen Bestandserhaltungsprogramms „Metamorfoze“ des Ministeriums für Bildung, Kultur und Wissenschaft erfolgte. Insgesamt wurden hierbei 350.000 Seiten größerer niederländischer Zeitungen digitalisiert:⁴³ zunächst drei Zeitungen aus der Zeit zwischen 1920 und 1929 im Rahmen des Projekts „Roaring Twenties“, später drei Zeitungen aus der Zeit von 1910 bis 1919 (eine linksorientierte, eine liberale und eine dem Zentrum nahestehende Zeitung). Von Letzteren wurden zunächst Mikrofilme erstellt, die als Vorlage für die Digitalisierung dienten.⁴⁴

Bereits früh im Prozess der Projektplanung entschied man sich, die Digitalisierungsarbeiten an externe Dienstleister zu vergeben. Hierzu wurden umfassende Marktstudien und Umfragen zum Stand der Technologie in der Zeitungsdigitalisierung, zum Zusammenhang zwischen verschiedenen Typen von Mikrofilm, Digitalisierung und OCR sowie schließlich zu verschiedenen Dateiformaten durchgeführt.⁴⁵ Aufbauend auf den Ergebnissen dieser Studien und einer europaweiten Ausschreibung wurde die Hamburger Firma CCS mit der Durchführung der Digitalisierung beauftragt.

Welche Rechercheoptionen sind für die Digitalisate vorgesehen? Aus den digitalisierten Zeitungsseiten werden die Volltexte extrahiert und die Ausgaben auf Articlebene erschlossen. Für die Recherche werden zunächst die Suche in den Volltexten, die Suche nach Zeitungstitel und Datum, eine Suche in den Titeln sowie ein Browsing zur Verfügung stehen. Perspektivisch sollen diese Optionen erweitert werden und unter anderem eine Suche nach Artikelkategorien, eine Eingrenzung nach Geographica sowie verschiedene Thesauri, beispielsweise für Namen, umfassen.⁴⁶

Über diese Maßnahmen der Königlichen Bibliothek hinaus werden in den Niederlanden in mehr als 40 Institutionen rund 150 verschiedene Zeitungstitel digitalisiert.⁴⁷ Datenformate, Erschließungsart und -tiefe sowie Recherchemöglichkeiten sind dabei höchst unterschiedlich. Ausgehend von dieser heterogenen Situation plant die Königliche Bibliothek, in Abstimmung mit den digitalisierenden Institutionen Verfahren und Standards zu entwickeln, um die verfügbaren digitalisierten historischen Zeitungen über ein gemeinsames Portal recherchierbar und nutzbar zu machen.⁴⁸

⁴³ Diese sind verfügbar über die Website „Kranten in Beeld“ (<http://kranten.kb.nl/index.html>) und sollen letztendlich in die Datenbank integriert werden.

⁴⁴ Vgl. <http://www.kb.nl/hrd/digitalisering/archief/roaringtwenties-en.html>; <http://www.kb.nl/hrd/digitalisering/archief/oorlog-en-revolutie-en.html>.

⁴⁵ Vgl. Edwin Klijn: The Current State-of-the-Art in Newspaper Digitization. A Market Perspective, in: D-Lib Magazine 14 (2008) 1/2, <http://www.dlib.org/dlib/january08/klijn/01klijn.html>; Verheusen: Newspaper Digitisation in the Netherlands, S. 58-59.

⁴⁶ Vgl. <http://www.kb.nl/hrd/digi/ddd/zoeken-en.html>.

⁴⁷ Als Beispiel für ein solches Projekt kann hier dienen: Eddy van der Noord: A Dutch Local Newspaper Digitised and Online in a Public-Private Partnership, in: LIBER Quarterly 18 (2008) 3/4, S. 394-398.

⁴⁸ Vgl. Verheusen: Newspaper Digitisation in the Netherlands, S. 61.

Hierzu konstituierte sich auf Initiative der Königlichen Bibliothek und „Digitaal Erfgoed Nederland“ Ende 2006 eine „Nationale werkgroep krantendigitalisering“.⁴⁹ Diese Expertengruppe veröffentlichte am 7.11.2008 das „Manifest Krantendigitalisering“. Darin wurde auf die Bedeutung von Zeitungen und ihrer Digitalisierung hingewiesen, das Engagement des Staates für eine nationale Infrastruktur für digitales Erbe betont und erörtert, dass Zeitungsdigitalisierung hierzu ein wichtiger Beitrag sei. Angesichts zahlreicher Digitalisierungsprojekte und des hohen Aufwandes sowie hoher Kosten für die Zeitungsdigitalisierung komme es darauf an, dass alle digitalisierenden Institutionen ihr Wissen miteinander teilten und ihre Projekte einander bekannt machten; das schließe auch die Beteiligung von Rechteinhabern, Finanziers und Softwareentwicklern ein. Ein wesentlicher Teil einer solchen koordinierten Anstrengung sei aber ein nationales Portal für digitalisierte Zeitungen.⁵⁰ Als erster Schritt in diese Richtung wurde von der Königlichen Bibliothek zunächst eine detaillierte Übersicht über die digitalisierten Bestände erarbeitet, um (Doppel-) Digitalisierungen bereits elektronisch vorliegender Bestände vermeiden zu können.⁵¹ Eine annähernd vollständige Übersicht ist im Internet verfügbar; Institutionen, die Zeitungsdigitalisierungsprojekte planen, sollen ihre Projektdaten dort angeben.⁵² Es wird abzuwarten sein, in welchem Umfang die angestrebte Vereinheitlichung gelingt und ob die „Databank Digitale Dagbladen“ den Ausgangspunkt einer langfristigen Entwicklung darstellt oder letzten Endes lediglich ein singuläres Projekt bleibt.

c. Großbritannien

Auch in Großbritannien sind (historische) Zeitungen in zahlreichen Gedächtnisinstitutionen in unterschiedlichem Umfang vorhanden. Anders als in Deutschland oder den Niederlanden nimmt jedoch die British Library (BL) als Nationalbibliothek eine dominierende Stellung ein. Sie besitzt eine beinahe vollständige Sammlung britischer und irischer Zeitungen seit 1840, sie sammelt systematisch und möglichst vollständig alle britischen und irischen Zeitungen, sowohl im Rahmen von Pflichtabgaben als auch durch Kauf und Tausch.⁵³ Seit 1869 gilt die Pflichtabgabe, auch für Zeitungen. Die Londoner Ausgaben der nationalen Tages- und Sonntagszeitungen seit 1801 sind vollständig, die britischen nationalen und regionalen Zeitungen seit dem 19. Jahrhundert sind beinahe

⁴⁹ Vgl. <http://www.den.nl/kennis/thema/kranten>.

⁵⁰ Vgl. Nationale Werkgroep Krantendigitalisering: Manifest Samenwerking Krantendigitalisering, 2008.

⁵¹ Darüber hinaus wurden mit mehreren Institutionen Abkommen geschlossen, um deren Bestände im Rahmen der Datenbank der Königlichen Bibliothek zugänglich zu machen.

⁵² Vgl. http://matrix.den.nl/matrix.aspx?matrixid=krantendigitalisering&view=Digitaal_Erfgoed; <http://www.kb.nl/hrd/digi/ddd/links-en.html>.

⁵³ Vgl. Edmund King: Digitisation of Newspapers at the British Library, in: The Serials Librarian 49 (2005) 1/2, S. 165-181.

komplett vorhanden, Zeitungen vor 1801 in unterschiedlichem Vollständigkeitsgrad, wobei die ältesten vorhandenen Zeitungen auf das Jahr 1699 datieren.⁵⁴ Die Bestände umfassen insgesamt rund 660.000 Bände, 370.000 Mikrofilmrollen sowie schließlich elektronische Datenträger.⁵⁵ Diese dominierende Stellung der British Library hinsichtlich des Bestandes an Zeitungen wurde durch die landesweiten Bestandserhebungs- und Verfilmungsprogramme NEWSPLAN und NEWSPLAN 2000 bestätigt.⁵⁶ Im Rahmen von NEWSPLAN wurden seit 1980 in den zehn bibliothekarischen Regionen des Landes möglichst komplette Erscheinungsverläufe von (regionalen und lokalen) Zeitungen im Vereinigten Königreich erfasst, erschlossen und so weit wie möglich mikroverfilmt und konservatorisch behandelt. NEWSPLAN 2000 zielte auf die Mikroverfilmung von rund 1.300 Titeln.

Die British Library kann als nationales Kompetenzzentrum für die Digitalisierung von Bibliotheksbeständen im Allgemeinen und die Zeitungsdigitalisierung im Besonderen gelten. Insbesondere in Hinblick auf Letztere nimmt sie eine dominierende Stellung ein, da abgesehen von wenigen Ausnahmen an anderen Institutionen⁵⁷ sämtliche Zeitungsdigitalisierungsprojekte in Großbritannien durch die British Library durchgeführt werden. Doch wie verfährt eine Institution, bei der sowohl das gesamte Quellmaterial als auch die Kompetenz liegt, in deren Verantwortung eine Digitalisierung also vollständig fällt?

Zunächst wurden in einem ersten Projekt 1992 die Chancen, Risiken und spezifischen Problemfelder einer (Zeitungs-)Digitalisierung vom Mikrofilm ermittelt. Hierzu wählten die Projektverantwortlichen die auf Mikrofilm vorliegende „Burney Collection of Newspapers“, die zahlreiche britische und irische Zeitungen von der Mitte des 17. bis zum Ende des 18. Jahrhunderts beinhaltet. Sie bildete ein ideales Objekt für das Projekt, da zum einen die Mikrofilme nur in schlechter Qualität vorlagen, sie allerdings gleichzeitig eine der meistbenutzten Mikrofilmsammlungen der Bibliothek darstellte. Insgesamt umfassten die vorliegenden Mikrofilme rund 650.000 Bilder.

In einem ersten Schritt wurden Zeitungen vom Ende des 18. Jahrhunderts digitalisiert – jedoch nicht vom Master-Mikrofilm, da Befürchtungen bestanden, diesen hierbei zu beschädigen. Da aber die Arbeitskopien von (noch) schlechterer Qualität waren, mussten die Scan-Parameter für beinahe jedes Bild neu eingestellt werden, was zu einem extrem

⁵⁴ Vgl. <http://www.bl.uk/reshelp/findhelprestype/news/blnewscoll/>.

⁵⁵ Insgesamt rund 52.000 Titel, wobei hierzu allerdings auch „magazines“ und „periodicals“ gerechnet werden.

⁵⁶ Vgl. <http://www.newsplan.co.uk>; <http://news.vispweb3.userarea.co.uk/index.html>.

⁵⁷ Für ein solches Projekt vgl. Suzanne Mieczkowska/Kathryn Pryor: Digitised Newspapers at Norfolk and Norwich Millennium Library, in: Collection Building 21 (2002) 4, S. 155-160. Daneben haben einige Zeitungsverlage ihre Archive digitalisiert und über das Internet öffentlich zugänglich gemacht.

verlangsamten Prozess führte. Mit der technischen Weiterentwicklung verbesserten sich die Ergebnisse, so dass die Zeitungen inzwischen digitalisiert und erschlossen vorliegen.⁵⁸

Nach diesen ersten Erfahrungen und dem Fortschritt in der Digitalisierungstechnik erfolgte 2000/2001 die Planung für den eigentlichen Einstieg in die Massendigitalisierung von historischen Zeitungen. Dieser sollte im Rahmen des Digitalisierungsprogramms der britischen Forschungsförderungsorganisation, des „Joint Information Systems Committee“ (JISC) erfolgen und durch dieses finanziert werden.⁵⁹

In Vorbereitung auf dieses Projekt wurde zunächst Anfang 2001 zusammen mit Partnern ein Pilotprojekt initiiert, in dem ein System für die Digitalisierung, Indexierung und Präsentation der historischen Zeitungen der British Library entwickelt und grundlegende Fragen zum Ablauf eines solchen Digitalisierungsprojekts geklärt wurden.⁶⁰ Insbesondere wurde getestet, eine kritische Masse mikroverfilmter Zeitungen in relativ kurzer Zeit zu digitalisieren, die digitalisierten Dokumente automatisch zu indexieren und sie mit fortgeschrittenen Suchmöglichkeiten einem möglichst breiten Publikum online bereitzustellen.

Das Projektteam entschied, lediglich gemeinfreie Zeitungen auszuwählen, um den Zeitplan nicht durch langwierige Verhandlungen mit Verlagen und Rechteinhabern zu verzögern. Es wurden mehrere für die britische Geschichte bedeutende Jahre bestimmt und aus diesen jeweils für bestimmte Monate komplette Erscheinungsverläufe ausgewählter Titel digitalisiert. Es wurden somit mehrere weit auseinander liegende Zeiten und unterschiedliche Zeitungen bearbeitet, so dass Erfahrungen mit einer großen Bandbreite an Formaten und Spezifika verschiedener Zeitungen gesammelt werden konnten. Die eigentliche Digitalisierung wurde an externe Dienstleister vergeben: Gescannt wurde durch die OCLC Preservation Resources, die weitere Bearbeitung erfolgte mit der ActivePaper Software von Olive.

Aufbauend auf den Erfahrungen aus der Digitalisierung der „Burney Collection“ sowie dem „Newspaper Pilot“ begann 2004 das eigentliche Projekt der „19th-Century British

⁵⁸ Graham Jefcoate: The Digitisation of the Burney Collection of Early Newspapers at the British Library, in: Hartmut Walravens/Edmund King (Hrsg.): Newspapers in International Librarianship. Papers presented by the Newspaper Section at IFLA General Conferences, München 2003 (IFLA Publications, Bd. 107), S. 183-187; Edmund King: Digital Historic Newspapers Online: Prospects and Challenges, in: Hartmut Walravens (Hrsg.): Newspapers Collection Management: Printed and Digital Challenges. Proceedings of the International Newspaper Conference, Santiago de Chile, April 3-5, 2007, München 2008 (IFLA Publications, Bd. 133), S. 59-78, hier S. 68-69.

⁵⁹ Vgl. JISC Digitisation Strategy, Februar 2008, http://www.jisc.ac.uk/media/documents/programmes/digitisation/jisc_digitisation_strategy_2008.doc.

⁶⁰ King: Digitisation of Newspapers, S. 8.

Newspapers”⁶¹. Vor dem Hintergrund des Vergleichs mit den anderen Staaten ergeben sich hierbei folgende Fragen: Welche Ziele wurden mit diesem Projekt verfolgt? Wie erfolgte die Auswahl der digitalisierten Zeitungen? Wie wurde der Prozess der Digitalisierung organisiert? Und schließlich, wie werden die Digitalisate präsentiert?

Als Ziel des Projekts wurde offiziell angegeben, es sollte das gesamte Spektrum britischer Zeitungen abdecken und damit den freien Zugang zu solchen Materialien ermöglichen, die vielfach bisher nur über den Londoner Lesesaal zugänglich waren. Zugleich sollten fortgeschrittene und komplexere Suchmöglichkeiten implementiert werden, wie beispielsweise die Volltextsuche oder die parallele Suche in mehreren Zeitungen. Hiermit zielte das Projekt im Übrigen genau auf die in der Einleitung genannten Vorteile der Digitalisierung.

Um diese Ziele zu erreichen, war ein breit angelegtes Sample vorgesehen, dessen Zusammensetzung in einem mehrstufigen Auswahlprozess erfolgte und das letztlich sowohl in London erschienene Tageszeitungen von nationaler Reichweite, englische, schottische, walisische und nordirische regionale Zeitungen als auch Zeitungen zu bestimmten Themen wie dem Viktorianischen Radikalismus oder dem Chartismus umfasste. Für die eigentliche Auswahl wurde dann eine Reihe von Entscheidungskriterien erarbeitet:

- möglichst komplette Erscheinungsverläufe
- möglichst komplette Abdeckung, zeitlich des 19. Jahrhunderts sowie geographisch des Vereinigten Königreichs
- eine Kombination von bekannteren und weniger bekannten Titeln
- Berücksichtigung der Verbreitung und der Auflagenhöhe einzelner Titel sowie unterschiedlicher Ausgaben einer Zeitung
- Berücksichtigung eines speziellen Inhaltes, wie beispielsweise einer deutlichen politischen Ausrichtung sowie schließlich die Frage, ob die Zeitung meinungsbildend war.⁶²

Außerdem sollten nur Titel gescannt werden, für die kein Urheberrechtsschutz mehr bestand – was in der Praxis bedeutete, dass nur Titel ausgewählt wurden, die älter als 100 Jahre waren. Darüber hinaus führte das Projektteam weitere Untersuchungen und Verhandlungen, um die urheberrechtliche Situation so präzise wie möglich zu klären.

Die Auswahl erfolgte gemeinsam durch den Projektstab und ein nicht detailliert spezifiziertes „User Panel of Experts“. Aus der ursprünglich vom Projektstab erstellten

⁶¹ King: Digitisation of Newspapers, S. 166-181; Jane Shaw: 10 Billion Words: The British Library British Newspapers 1800-1900 Project. Some Guidelines for Large-Scale Newspaper Digitisation, in: Hartmut Walravens (Hrsg.): International Newspaper Librarianship for the 21st Century, München 2006 (IFLA Publications, Bd. 118), S. 27-43.

⁶² Vgl. http://newspapers.bl.uk/blcs/page.do?page=/researchguide.jsp&docPage=blcs_24.htm.

Liste von mehr als 160 Titeln wurden 48 ausgewählt, darunter 17 „nationale“ und 29 regionale Titel.⁶³ Über diese Titelauswahl wurde dann mittels eines Online-Fragebogens die wissenschaftliche „Community“ befragt, wobei sich vor allem Bibliothekare und in geringerer Zahl Forscher, Lehrer und Studenten beteiligten. Diese stimmten dem Projektansatz und der Titelauswahl grundsätzlich zu, bemängelten jedoch das Fehlen von irischen Titeln, so dass dem Sample eine Zeitung aus Dublin hinzugefügt wurde.⁶⁴ Die Auswahl umfasst somit 49 Titel mit insgesamt rund 2 Millionen Zeitungsseiten – eine im Hinblick auf die Seitenzahl umfangreiche, verglichen mit den mehreren (zehn)tausend im Vereinigten Königreich publizierten Titeln zunächst jedoch kleine Auswahl.

Ähnlich wie im Projekt des Newspaper Pilot wurde auch in diesem Projekt eine klare Aufgaben- und Arbeitsteilung vorgenommen. Materialauswahl, Konservierung sowie die Mikroverfilmung der Zeitungen erfolgten in der British Library durch Mitarbeiter der Bibliothek, die eigentliche Digitalisierung, die Inhaltsextraktion und die Erstellung des Präsentationssystems durch Dienstleister, die im Rahmen einer europaweiten Ausschreibung gewonnen wurden.

Der Workflow umfasste das Scannen – vorwiegend von Mikrofilm sowie ggf. die Erstellung neuer Mikrofilme –, eine umfangreiche Erschließung inklusive OCR mit einer intendierten Genauigkeit von mindestens 80 Prozent Artikelseparierung und -erschließung, das Generieren umfangreicher Metadaten sowie die Entwicklung einer intuitiv zu bedienenden Oberfläche mit umfangreichen Recherchemöglichkeiten.⁶⁵

Auf der Portalseite „British Newspapers 1800-1900“ sind sämtliche digitalisierten Zeitungen recherchierbar.⁶⁶ Für Bildungseinrichtungen im Vereinigten Königreich („Higher Education Organisations“) ist der Zugang kostenlos, alle anderen Nutzer können gegen eine Gebühr den Zugriff auf eine definierte Anzahl von Artikeln erwerben.⁶⁷ Die Weboberfläche bietet zum einen die Möglichkeit, nach dem Einstieg über Titel oder Orten durch die Zeitungen zu „blättern“. Zum anderen können Nutzer nach Wörtern aus dem Volltext, nach Daten, Titeln, Zeitungstiteln oder Wochentagen suchen. Diese Suche kann

⁶³ Vgl. Edmund King: 19th Century British Library Newspapers: Utilising the Online Database, in: Hartmut Walravens (Hrsg.): The Impact of Digital Technology on Contemporary and Historic Newspapers. Proceedings of the International Newspaper Conference, Singapore, 1.-3. April 2008 and papers from the IFLA World Library and Information Congress, Québec, Canada, August, 2008, München 2008 (IFLA Publications, Bd. 135), S. 47-54, hier S. 48; http://newspapers.bl.uk/blcs/page.do?page=/researchguide.jsp&docPage=blcs_24.htm.

⁶⁴ Vgl. King: Digital Historic Newspapers Online, S. 61.

⁶⁵ Vgl. Kapitel B.2 zu grundsätzlichen Aspekten der einzelnen Workflow-Schritte.

⁶⁶ Vgl. <http://newspapers.bl.uk/blcs/>.

⁶⁷ Abgesehen vom kostenlosen Zugriff auf die digitalisierten Ausgaben des Penny Illustrated Paper und von The Graphic.

über einen oder mehrere Titel erfolgen und unter anderem nach Datum, Artikelkategorie⁶⁸ und Erscheinungsfrequenz eingegrenzt werden. Suchanfragen lassen sich speichern, Dokumente und Texte herunterladen. Neben der eigentlichen Suche bietet das Portal Einführungen in die britische Zeitungsgeschichte des 19. Jahrhunderts, in den historischen Kontext – das heißt zu verschiedenen Themenbereichen der britischen Geschichte –, zu ausgewählten Persönlichkeiten sowie detaillierte Informationen über das Projekt selbst. Schließlich werden kurze Erläuterungen zu den einzelnen Zeitungstiteln geboten. Bereits an dieser Stelle ist festzuhalten, dass hier im Unterschied beispielsweise zur digitalen Zeitungsdatenbank in den Niederlanden oder auch den größeren deutschen Projekten umfangreiche Ergebnisse vorliegen und online recherchierbar sind.

Im Rahmen der zweiten Phase des JISC-Programms zur Digitalisierung wurde das Anschlussprojekt „British Newspapers 1620-1900“ aufgelegt, innerhalb dessen rund 1,25 Millionen Zeitungsseiten digitalisiert werden sollen.⁶⁹

Bis April/Mai 2009 wurden Zeitungsdigitalisierungsprojekte im Vereinigten Königreich (fast) ausschließlich durch die British Library durchgeführt. Ende April 2009 gab allerdings die Walisische Regionalregierung bekannt, der Walisischen Nationalbibliothek 2 Millionen Pfund bereitzustellen, mit der rund 300 Zeitungen und Zeitschriften digitalisiert werden sollen – im Rahmen eines Programms, das einen umfassenden Teil des gedruckten kulturellen Erbes Wales’ online zur Verfügung stellen will.⁷⁰ Informationen zu Details des Programms sind noch nicht veröffentlicht worden; es ist insbesondere unklar, ob die gleichen technischen Standards und Formate wie durch die British Library eingesetzt und wie die Ergebnisse recherchierbar sein werden; das heißt auch, ob gegebenenfalls eine gemeinsame Rechercheoberfläche existieren wird.

d. USA

Ein Vergleich der Zeitungsdigitalisierung in den USA mit den deutschen Strukturen ist besonders interessant, da in mehrfacher Hinsicht parallele Ausgangsbedingungen herrschen. Die USA sind wie Deutschland stark föderal organisiert, mit wesentlichen Kompetenzen der Bundesstaaten auch in der Kultur- und Bildungspolitik. Auch die Bibliothekslandschaft ist nicht so monolithisch oder zentralisiert wie beispielsweise in den Niederlanden oder in Großbritannien, es existiert eine Vielzahl leistungsfähiger und (inter)national bedeutender

⁶⁸ Nachrichten, Werbung, Wirtschaftsnachrichten, „Arts and Sports“ und „People“. Vgl. dazu auch die ursprünglich vorgesehenen Kategorien – King: Digital Historic Newspapers Online, S. 66.

⁶⁹ Vgl. <http://www.jisc.ac.uk/whatwedo/programmes/digitisation/newspapers2.aspx>; http://www.jisc.ac.uk/media/documents/programmes/digitisation/bl_newspapers_public_plan.pdf.

⁷⁰ Vgl. <http://digitisation.jiscinvolve.org/2009/05/01/more-digitisation-funding-for-the-national-library-of-wales/>.

Bibliotheken. Auf der anderen Seite sind, im Hinblick auf die hier interessierende Fragestellung, einige wesentliche Unterschiede zwischen den USA und Deutschland zu benennen: Die „Library of Congress“ (LC) besitzt eine Stellung, Ressourcen und Kompetenzen, die über die der deutschen Nationalbibliothek hinausreichen. Daneben existiert mit dem „National Endowment for the Humanities“ (NEH) eine Förderungseinrichtung, die speziell die „humanities“, also Geisteswissenschaften fördert, und zwar nicht nur genuine Forschung, sondern auch Infrastruktur und die Bewahrung des „kulturellen Erbes“ im weiteren Sinne. Und schließlich sind – im Unterschied zu Deutschland – die in den USA vorhandenen Zeitungsbestände aufgrund eines landesweiten Erschließungsprogramms weitestgehend nachgewiesen und erschlossen: Das von 1982 bis 2007 durchgeführte „United States Newspaper Program“ (USNP)⁷¹ zielte darauf ab, in den gesamten USA der Bestand an Zeitungen von 1690 bis in die Gegenwart ausfindig zu machen, zu katalogisieren sowie in größerem Umfang mikrozuverfilmen. Das Programm wurde gemeinsam von der Bundesregierung und den Staaten und Territorien realisiert, wobei das NEH für die Finanzierung und die Library of Congress für die technische Unterstützung zuständig war. Die Erschließungsprojekte wurden dezentral in den Staaten und Territorien durchgeführt, wobei jeweils die Einrichtung mit dem größten Zeitungsbestand federführend. Die Daten jeder Einrichtung wurden an CONSER („Cooperative Online Serials Cataloging Program“) beziehungsweise die Newspaper Union List gemeldet.

Damit sind die für den in dieser Arbeit vorzunehmenden Vergleich wesentlichen Gemeinsamkeiten und Unterschiede der Rahmenbedingungen genannt. Die in diesen Kontexten gefundenen Lösungen und Strukturen für die Digitalisierung historischer Zeitungen sollen im Folgenden betrachtet werden. Grundsätzlich finden sich in den USA Projekte zur Zeitungsdigitalisierung auf mehreren Ebenen, das heißt zum einen regional, zum anderen auf der Ebene einzelner Bundesstaaten und schließlich landesweit. Vor dem Hintergrund, eine langfristige Perspektive auch für Deutschland aufzeigen zu wollen, interessiert hier das landesweite Zeitungsdigitalisierungsprogramm besonders. Um dessen Aufbau nachvollziehbar zu machen, sollen im Folgenden seine wesentlichen Merkmale umrissen werden:

Durch das „National Digital Newspaper Program“ (NDNP) soll, so die verlautbarte Absicht, langfristig eine umfassende Auswahl der „historisch bedeutenden“ Zeitungen aller Bundesstaaten und Territorien der USA digital verfügbar gemacht werden, die zwischen 1836 und 1922 veröffentlicht wurden.⁷² Dieser Zeitraum wurde gewählt, da die Zeit vor 1836 durch andere (Digitalisierungs-)Programme erschlossen und die Zeit nach 1922

⁷¹ Vgl. <http://www.neh.gov/projects/usnp.html>.

⁷² Vgl. <http://www.neh.gov/projects/ndnp.html>.

urheberrechtlich problematisch ist – vor 1922 veröffentlichte Publikationen gelten in den USA hingegen als „public domain“.⁷³

Das Konzept des Programms sieht eine Steuerung, anders als in den Niederlanden oder in Großbritannien, nicht durch eine zentralisierte Digitalisierung in einer einzigen Bibliothek oder durch einen übergeordneten Expertenrat, sondern vielmehr durch das Instrument einer zentralen Finanzierung vor. Die Digitalisierung erfolgt dezentral in den einzelnen Bundesstaaten, finanziert wird sie jedoch durch das NEH⁷⁴ und technisch betreut durch die Library of Congress⁷⁵. Beide Institutionen haben dadurch die Möglichkeit, über die Bedingungen für eine Förderung eine (indirekte) Steuerung vorzunehmen. Dieses Vorgehen trägt der föderalen Struktur des Landes ebenso Rechnung wie dem Umstand, dass nicht alle Zeitungen in einer einzelnen Institution zentral vorhanden sind.

Die technischen Spezifikationen und Vorgaben für die Digitalisierung wurden von der Library of Congress erarbeitet.⁷⁶ Ohne hier auf einzelne Elemente eingehen zu können, kann gesagt werden, dass dabei in verschiedener Hinsicht der Gedanke einer „Offenheit“ verfolgt wurde. Dies beinhaltete im engeren Sinne die persistente Adressierung der Digitalisate sowie die Verwendung von Open-Source-Software und offenen technischen Standards; darüber hinaus sollte aber auch die allgemeine und freie Zugänglichkeit der Ergebnisse des Programms sichergestellt sein.⁷⁷ Im Rahmen dieser Vorgaben haben die einzelnen Institutionen die Möglichkeit, den Digitalisierungsprozess weitgehend selbst auszugestalten. Sie können entscheiden, ob sie die Digitalisierung an Dienstleister vergeben (wie zum Beispiel in Kalifornien) oder im Hause vornehmen (so in Kentucky).

In jedem Bundesstaat beziehungsweise Territorium wird eine Institution gefördert, die die Digitalisierung durchführt oder koordiniert. Dies ist im Regelfall die Institution mit dem größten Zeitungsbestand, die dann wiederum mit anderen Einrichtungen kooperieren kann, um die gewünschte Bandbreite der „historically significant newspapers“ zu erreichen. Für jeden Staat beziehungsweise Antragsteller sind Mittel für die Digitalisierung von 100.000 Seiten pro Zeitschnitt vorgesehen.⁷⁸

⁷³ Vgl. <http://www.oah.org/pubs/nl/2004may/cole.html>; Helen Aguera: The U.S. National Digital Newspaper Program: Thinking Ahead, Designing Now, Part I: Program Development, in: Hartmut Walravens (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 79-81, hier S. 80.

⁷⁴ Vgl. <http://www.neh.gov/whoweare/index.html>; Aguera: U.S. National Digital Newspaper Program, S. 79.

⁷⁵ Zugleich plant die Library of Congress, selbst einen größeren Teil ihres eigenen Zeitungsbestands zu digitalisieren und einzuspielen; vgl. <http://www.neh.gov/projects/ndnp.html>.

⁷⁶ Vgl. Library of Congress: The National Digital Newspaper Program (NDNP) Technical Guidelines for Applicants, Washington, D.C. 2009.

⁷⁷ Mark Sweeney: The U.S. National Digital Newspaper Program: Thinking Ahead, Designing Now, Part II: Preservation Planning, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 83-85.

⁷⁸ Insgesamt standen ursprünglich Mittel für rund 30 Millionen Seiten zur Verfügung.

Angesichts eines Vorhabens dieses Umfangs stellen sich mehrere Fragen: Wie erfolgt die Auswahl der zu digitalisierenden Zeitungen? Denn es ist explizit nicht vorgesehen, sämtliche Zeitungen des Landes zu digitalisieren. Wie ist das Programm in zeitlicher Hinsicht strukturiert? Wie erfolgt die Finanzierung der Einzelprojekte? Und schließlich, wie werden die Digitalisate zusammengeführt und präsentiert? Diese Fragen sollen nachfolgend erörtert werden.

Das NDNP empfiehlt den Institutionen in den einzelnen Staaten folgendes Vorgehen für die Auswahl der Zeitungstitel: Es sollte ein beratendes Gremium eingerichtet werden, dem unter anderem Wissenschaftler, Lehrende, Bibliothekare und Archivare angehören. Dem beratenden Gremium sollten Informationen über die einzelnen Titel genannt werden, damit es eine breite, in folgenden Aspekten repräsentative Auswahl treffen kann: nach Leserschaft, politischer Orientierung, geographischer Verteilung. Die ausgewählten Zeitungstitel sollten die politische, wirtschaftliche und kulturelle Geschichte des Staates widerspiegeln; daher kommen vor allem Zeitungen mit staatsweitem oder regionalem Einfluss in Betracht und insbesondere solche Titel, die im Staat oder einzelnen Regionen bzw. Counties als „paper of record“ gelten und die auch offizielle Ankündigungen und Mitteilungen beinhalten. Darüber hinaus sollen möglichst alle größeren Bevölkerungszentren abgedeckt sein. Bevorzugt werden Zeitungen mit langem kontinuierlichen Erscheinungsverlauf, um in weiteren Phasen des NDNP an vorausgegangene Projekte anknüpfen zu können. Dabei sollen speziell die Zeitungen beachtet werden, die ihr Erscheinen eingestellt und/oder keinen Eigentümer mehr haben, so dass eine Digitalisierung in anderem Zusammenhang, durch eine andere Institution unwahrscheinlich erscheint.⁷⁹

Das NDNP ist aus verschiedenen Gründen in mehrere Phasen eingeteilt. Einerseits sollten in einer ersten Phase mit einer geringen Zahl von geförderten Bibliotheken Erfahrungen gesammelt werden, um möglicherweise technische Vorgaben und Parameter modifizieren zu können. Vor allem aber ermöglicht die Einteilung in mehrere Phasen eine Entzerrung der Finanzierung. In jeder Phase soll sowohl die Zahl der geförderten Institutionen als auch der zu bearbeitende Zeitraum weiter gefasst werden. Es ist geplant, jedes Jahr eine weitere Ausschreibung für Förderungen vorzunehmen, bis alle Staaten und Territorien eine Auswahl ihrer Zeitungen digitalisiert und zur Verfügung gestellt haben.⁸⁰ In der von 2005 bis 2007 laufenden ersten Phase wurde sechs Bibliotheken die Digitalisierung für den Zeitraum von 1900 bis 1910 bewilligt, Anfang 2007 wurden die ersten rund 226.000 Seiten

⁷⁹ Vgl. <http://www.neh.gov/grants/guidelines/ndnp.html>; <http://www.loc.gov/ndnp/content-selection.html>. Zu fragen wäre, inwiefern auch nachgewiesen muss, dass die betreffende Zeitung nicht durch einen der noch zu behandelnden kommerziellen Anbieter digitalisiert worden ist.

⁸⁰ Vgl. Aguera: U.S. National Digital Newspaper Program, S. 81.

online gestellt. Ab 2007 und 2008 wurden Projekte in neun weiteren Staaten gefördert und der Zeitraum (auch für die bereits vorher geförderten Projekte) auf 1880 bis 1910 beziehungsweise 1922 ausgeweitet.⁸¹ Die nächste Phase schließlich beginnt 2009 und soll Zeitungen der Erscheinungsjahre 1860 bis 1922 umfassen; dabei werden fünfzehn Einrichtungen beziehungsweise Staaten gefördert.⁸²

Die Förderung erfolgt als Kooperationsvereinbarung zwischen dem NEH und der jeweiligen geförderten Institution mit einer Laufzeit von jeweils zwei Jahren und umfasst eine Förderungssumme von bis zu 400.000 Dollar pro Staat/Institution. Das NEH vereinbart dabei mit den Empfängern die endgültige Auswahl der Titel, Eckdaten des Prozesses sowie die Evaluation der technischen Spezifikation; Mitarbeiter der Library of Congress überprüfen die Übereinstimmung der generierten Daten mit den technischen Vorgaben.⁸³ Die Vereinbarung enthält darüber hinaus einen von beiden Seiten genehmigten Arbeitsplan, der neben der Auswahl der zu digitalisierenden Zeitungstitel Angaben zu den einzelnen Schritten des Workflows umfasst: zur technischen Analyse des verfügbaren Mikrofilms; Details zum Scan und der Erstellung mehrerer Datenformate, der Volltextdateien sowie der notwendigen Metadaten und Dokumentation; zum Prozedere, wie die Images, Volltexte und Metadaten an die Library of Congress gemäß den technischen Richtlinien geliefert werden; sowie schließlich zum Zeitplan.⁸⁴

Die Digitalisate aller teilnehmenden Institutionen werden an die Library of Congress überspielt, hier gespeichert und in das dort entwickelte Portal „Chronicling America“ eingespielt.⁸⁵ Auf diesem Portal hat der Benutzer zum einen die Möglichkeit, das auf den Ergebnissen des USNP aufbauende „Newspaper Title Directory“⁸⁶ zu nutzen, um bibliographische und Bestandsangaben zu Zeitungen zu suchen. Zum anderen bieten sich ihm die bereits online gestellten, im Rahmen des NDNP digitalisierten Bestände zum Durchsuchen. Möglich ist dabei lediglich die Suche im Volltext (Stichworte oder Phrasen), die nach Datum, Zeitungen und Bundesstaaten eingegrenzt werden kann. Die Suchergebnisse, das heißt die Zeitungsseiten, können als PDF oder Jpeg2000 abgespeichert werden.

⁸¹ Vgl. <http://www.loc.gov/ndnp/listawardees.html>.

⁸² Vgl. <http://www.neh.gov/grants/guidelines/ndnp.html>; <http://www.loc.gov/ndnp/listawardees.html>.

⁸³ Die Überwachung der Dateiformate erfolgt dabei nach eigenen Angaben mit einem Java-basierten Tool, das für diesen speziellen Einsatzzweck über die Funktionalitäten von JHOVE hinausreicht. Vgl. Justin Littman/Dave Woodward: Technical Implementation Specifications, NDNP, Mai 2005, http://www.loc.gov/ndnp/pdf/NDNP_TechSpecs_0505.pdf.

⁸⁴ Vgl. <http://www.neh.gov/grants/guidelines/ndnp.html>.

⁸⁵ Vgl. George Schlukbier: The U.S. National Digital Newspaper Program: Thinking Ahead, Designing Now, Part IV: Repository Development, in: Hartmut Walravens (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 91-95.

⁸⁶ Rund 140.000 Titeleinträge und 900.000 Besitznachweise.

e. Kommerzielle Zeitungsdigitalisierungsprojekte

Bisher wurden mehrere von öffentlicher Seite initiierte Projekte betrachtet. Diese nehmen in den Diskussionen, die in der Bibliothekslandschaft geführt werden, einen breiten Raum ein. Daneben existieren jedoch auch kommerzielle Projekte, die im Vergleich zu den bisher betrachteten ein wesentlich größeres Ausmaß aufweisen und inzwischen eine große Auswirkung auch auf die staatlichen Initiativen haben. Um den internationalen Stand der Zeitungsdigitalisierung erfassen zu können, sind diese kommerziellen Projekte ebenfalls zu betrachten. Andernfalls könnten auch die Perspektiven der Zeitungsdigitalisierung nicht realistisch eingeschätzt und es könnte kein entsprechendes Konzept für Deutschland fundiert erarbeitet werden. Die folgende Betrachtung wird sich dabei auf die wesentlichen kommerziellen Zeitungsdigitalisierungsprojekte und Anbieter digitalisierter historischer Zeitungen konzentrieren. Insbesondere in den USA werden von einer Reihe von Firmen in großem Maße Zeitungen digitalisiert und angeboten. Die derzeit größten und auf dem Markt wichtigsten Anbieter sind ProQuest, Heritage Microfilms mit dem NewspaperARCHIVE sowie Readex. Schließlich ist auf die Aktivitäten von Google einzugehen, das inzwischen ebenfalls in die Zeitungsdigitalisierung eingestiegen ist.

ProQuest ist ein Anbieter von Inhalten und Datenbanken, der umfangreiche Archive mit Mikroformen besitzt, seit einigen Jahren aber auch digitale Inhalte anbietet, unter anderem die „Historical Newspapers“. Diese umfassen rund 34 Zeitungen von der Mitte des 18. Jahrhunderts bis in die Gegenwart, darunter unter anderem die New York Times und die Washington Post. Insgesamt sind rund 20 Millionen Seiten digitalisiert und durchsuchbar.⁸⁷ Das Angebot von ProQuest ist 2008 darüber hinaus durch den Kauf von Dialog, einem der weltweit führenden Anbieter von Datenbanken, insbesondere von digitalisierten historischen Zeitungen, erweitert worden.

Readex, zum Informationsanbieter NewsBank gehörig, bietet zahlreiche digitale Quellensammlungen an, darunter America's Historic Newspapers – nach eigenen Angaben rund 2.000 verschiedene Zeitungstitel zwischen 1690 und 1922 mit mehreren Millionen Seiten.⁸⁸

Heritage Microfilms schließlich, vor allem als Anbieter von Mikroformdienstleistungen aktiv, bietet mit dem NewspaperARCHIVE einen umfangreichen Bestand an digitalisierten historischen Zeitungen an, nach eigenen Angaben insgesamt rund 3.400 verschiedene Titel mit 25 Millionen Seiten.

Alle drei Anbieter bieten diese Inhalte über kostenpflichtige Datenbanken an. Während sich ProQuest und Readex dabei primär an Bibliotheken und andere Institutionen wenden,

⁸⁷ Vgl. <http://www.proquest.co.uk/en-UK/catalogs/databases/detail/pq-hist-news.shtml>.

⁸⁸ Vgl. <http://www.readex.com/readex/index.cfm?content=96>.

vermarktet NewspaperARCHIVE sein Angebot vor allem an Privatpersonen beziehungsweise Endnutzer.

Diese kommerziellen Anbieter verfügen bereits über digitalisierte Zeitungen in einem Umfang, an den die Projekte der Gedächtniseinrichtungen nicht heranreichen. Dennoch sind diese Unternehmen Teil eines etablierten Markts, insbesondere durch ihre Zielgruppen (Wissenschaftler, „private Forscher“), das Produktdesign (ausgefeilte Suchinstrumente) sowie die Vermarktung (Lizenzen oder Abonnement). Geradezu eine Revolutionierung auf diesem Markt der Zeitungsdigitalisierung und des Angebots digitalisierter historischer Zeitungen erfolgt derzeit durch Google. Durch mehrere Kooperationen und eigene Digitalisierungsprojekte ist Google in kurzer Zeit zum größten Anbieter digitalisierter Zeitungen geworden und stellt diese – bisher kostenpflichtigen Angebote – kostenfrei über das Internet zur Verfügung.⁸⁹

2006 hatte Google begonnen, in Kooperation mit der New York Times und der Washington Post deren bereits existierende digitale Archive zu indexieren und über Google News verfügbar zu machen. Im September 2008 verkündete Google, langfristig und umfassend historische Zeitungen zu digitalisieren und bereitstellen zu wollen. Dies sollte, ähnlich wie beim Google-Books-Projekt, wiederum in Kooperation mit den Bibliotheken geschehen, die ihre Bestände zur Verfügung stellen und diese als Gegenleistung kostenfrei digitalisiert bekommen sollen.⁹⁰

Darüber hinaus wurden Kooperationen mit den Firmen ProQuest und Heritage Microfilms geschlossen: ProQuest soll Google unter anderem sein Archiv an mikroverfilmten historischen Zeitungen (rund 10.000 Titel) zur Digitalisierung zur Verfügung stellen. Die Einzelheiten der Kooperation mit Heritage Microfilms sind noch nicht bekannt. Insbesondere ist unklar, inwieweit der Bestand von NewspaperARCHIVE nunmehr über Google kostenfrei zur Verfügung gestellt werden soll und ob Heritage seine Archive von rund 180.000 größtenteils noch nicht digitalisierten Rollen Mikrofilm mit historischen Zeitungen für die Digitalisierung zur Verfügung stellt.⁹¹ Im Dezember 2008 schließlich kaufte Google den Anbieter „Paper of Record“ und dessen Bestand an rund 20 Millionen digitalisierten Zeitungsseiten.⁹² Auch wenn die Bestände der ProQuest Historical Newspapers nicht über Google News Archive recherchierbar sein sollten, erhält Google über die genannten Kooperationen somit Zugriff auf mehr als 40 Millionen digitalisierte

⁸⁹ Vgl. auch: Thaddeus Lipinski: Google Digitisation Programme, in: Hartmut Walravens (Hrsg.): Newspapers Collection Management: Printed and Digital Challenges. Proceedings of the International Newspaper Conference, Santiago de Chile, April 3-5, 2007, München 2008 (IFLA Publications, Bd. 133), S. 101-104.

⁹⁰ Vgl. <http://googleblog.blogspot.com/2008/09/bringing-history-online-one-newspaper.html>.

⁹¹ Vgl. <http://www.heritagemicrofilm.com/Resources.aspx>.

⁹² Vgl. <http://hurstassociates.blogspot.com/2008/12/google-buys-papers-of-record.html>; <http://blog.searchenginewatch.com/081202-100005>.

Zeitungsseiten, auf Mikrofilm, der kostengünstig digitalisiert werden kann, sowie auf Know-how erfahrener Digitalisierungsdienstleister.⁹³

Die Aktivitäten von Google im Bereich der Digitalisierung haben auch Auswirkungen auf Digitalisierungsprojekte von Kultur- und Gedächtnisinstitutionen. So wird beispielsweise in der neuen Fassung der DFG-Praxisregeln verlangt, Antragsteller müssten überprüfen und belegen, dass ihr Digitalisierungsprojekt sich nicht mit kommerziellen Vorhaben überschneide. Explizit wird hierbei auf die Kooperation von Google mit der Bayerischen Staatsbibliothek verwiesen.⁹⁴ Diese Forderung gilt auch für jedes über diese Programmlinie geförderte Zeitungsdigitalisierungsprojekt. Dies könnte langfristig die Auswirkung haben, dass bestimmte Zeitungen nicht mehr im Rahmen öffentlich finanzierter Projekte digitalisiert werden, da ihre Digitalisierung durch Google erwartet wird. Im Hinblick auf die grundsätzliche Digitalisierung und Vermeidung doppelten Aufwands wäre dies zu begrüßen. Berücksichtigt werden müssen dabei jedoch auch die Erschließungstiefe und Rechercheoptionen einerseits bibliothekarischer Digitalisierungsprojekte, andererseits von Google. Google wird die Zeitungen über die Dienste Google News beziehungsweise Google News Archive zur Verfügung stellen. Die historischen Zeitungen werden parallel zu aktuellen Nachrichten recherchierbar sein. Die Recherchemöglichkeiten sind – derzeit – im Vergleich zu den Angeboten kommerzieller Anbieter oder von Bibliotheken eingeschränkt: Gesucht werden kann nach Stichwörtern und Phrasen, die Suche kann nach Datum, Sprache und Quelle (Zeitungstitel) eingegrenzt werden. Eine Suche nach Kategorien von Inhaltselementen, Personen oder Schlagworten ist nicht möglich. Darüber hinaus sind Funktionen wie eine Thumbnail-Übersicht oder Highlighting nicht implementiert.

Der (mögliche) Verzicht auf Digitalisierungsvorhaben durch die Bibliotheken scheint vor diesem Hintergrund zunächst eine Minderung der Recherchequalität zu bewirken. Gerade aus diesem Grund ist jedoch fraglich, ob dies eine logische und zu erwartende Entwicklung ist. Denn viele Forscher benötigen ausgefeiltere Suchinstrumente. Ob das andererseits zur Finanzierung entsprechender Digitalisierungsprojekte führt, ist wiederum ungewiss. Möglicherweise kommt es auch zur Koexistenz verschiedener Angebote – eines kostenpflichtigen Angebots mit umfangreichen Suchoptionen für Wissenschaftler sowie eines kostenfreien für Personen, die diese Optionen nicht benötigen. Auch wenn sich diese Fragen derzeit nicht beantworten lassen, so werden sie doch aufmerksam zu verfolgen sein. Abschließend kann allerdings festgehalten werden, dass die Initiative von Google

⁹³ Vgl. auch Astrid Verheusen: Mass Digitisation by Libraries: Issues Concerning Organisation, Quality and Efficiency, in: LIBER Quarterly 18 (2008) 1, S. 28-38, hier S. 28.

⁹⁴ Vgl. Deutsche Forschungsgemeinschaft: DFG-Praxisregeln „Digitalisierung“ zu den Förderprogrammen der Wissenschaftlichen Literatur- und Informationssysteme, Bonn 2009, S. 5.

Bewegung in ein Gebiet gebracht hat, in dem Bibliotheken gerade erst begonnen haben, über Versuchsstadien hinaus zu gelangen.

2. Stand der Technik

Neben den bisher betrachteten organisatorischen Fragen ist auch eine Reihe von technischen Aspekten für die Planung und insbesondere die Realisierbarkeit von Digitalisierungsprojekten relevant. Daher werden im Folgenden die wesentlichen technischen Fragen berücksichtigt und der Stand der Technik anhand einzelner Schritte eines idealtypischen Workflows⁹⁵ dargestellt, wie er sich durch die Erfahrungen aus der Digitalisierung von Bibliotheks- und Archivgut entwickelt hat. Konkret werden hier wichtige und kontroverse Teilaspekte der wesentlichen Digitalisierungsschritte Scan, Erschließung und Präsentation/Speicherung betrachtet. Der Fokus liegt dabei auf den richtungsweisenden Entscheidungen, die im Rahmen dieser Arbeitsschritte zu treffen sind und die sich merklich auf die zu kalkulierende Zeit und das Budget, mithin also die gesamte Planung auswirken. Da die vorliegende Arbeit nicht als Marktstudie konzipiert ist, werden konkrete Hardware- und Software-Produkte nicht dezidiert betrachtet.

Die technischen Entwicklungen sind, wie eingangs schon erwähnt, – in weit größerem Maße als organisatorische Fragen – international. Die nachfolgenden Ausführungen sollen daher nicht auf die oben betrachteten vier Staaten begrenzt bleiben, sondern berücksichtigen darüber hinaus auch Erfahrungen und Digitalisierungsprojekte in weiteren Ländern.

a. Scan

Der erste Schritt der Digitalisierung ist das Scannen. Hierbei sind eine Reihe von Schritten relevant und für die Qualität des Endprodukts ausschlaggebend: Wird vom Original oder einer Mikroform gescannt? In welcher Farbtiefe und Auflösung wird gescannt? Welche Dateiformate werden erzeugt?

Die erste Entscheidung bei der Digitalisierung von Zeitungen gilt der Frage, ob vom Papier-Original oder von der Mikroform gescannt werden soll. Während diese Frage für viele andere Medien irrelevant ist, stellt sie sich für Zeitungen in besonderem Maße.

⁹⁵ Der Begriff des „Workflow“ hat sich im Zusammenhang mit Digitalisierungsprojekten inzwischen auch im Deutschen durchgesetzt und bezeichnet die (gesamte) Prozessabfolge als Ganzes und Summe der Einzelschritte.

Spätestens seit Mitte des 20. Jahrhunderts sind Zeitungen in großem Umfang im Rahmen von Maßnahmen zur Bestandssicherung mikroverfilmt worden. Sie stehen also im Idealfall als Papieroriginal und als Mikrofilm zur Verfügung. Bei der Betrachtung aktuell laufender Projekte lässt sich allerdings kein Standard dahingehend erkennen, ob bevorzugt vom Mikrofilm oder vom original gescannt wird, wenngleich insgesamt dem Mikrofilm tendenziell der Vorzug gegeben zu werden scheint. Die für beide Vorgehensweisen bestehenden Vor- und Nachteile sollen im Folgenden umrissen werden.

Für die Verwendung von Mikrofilm als Quellmaterial wird vor allem angeführt, die Digitalisierung sei wesentlich günstiger und mit weniger Aufwand durchzuführen. So setzte die Bibliothek der University of Utah zunächst auf das Scannen vom Mikrofilm, da (in der Berechnung von 2003/2004) das Scannen einer Seite vom Mikrofilm mit 0,15 US-Dollar nur die Hälfte des Scans vom Original kosten würde. Als weiterer Vorteil wird die Verfügbarkeit genannt: Aufgrund der in mehreren Ländern durchgeführten, breit angelegten nationalen Programme zur Mikroverfilmung von Zeitungen sind von vielen Titeln komplette „Läufe“ vorhanden. Und da im Rahmen dieser Programme auch kollationiert und verteilte Bestände für die Verfilmung zusammengeführt wurden, kann eine zeitintensive Kollationierung im Vorfeld der Digitalisierung unterbleiben. Aus diesem Grund setzte beispielsweise die Australische Nationalbibliothek auf das Scannen vom Mikrofilm und auch die Vorgaben des NDNP sehen die Verwendung von Mikrofilm vor, da dieser ja durch das USNP vorliege.⁹⁶ Auch die Geschwindigkeit des Scan-Vorgangs wird vielfach hervorgehoben. Bereits beim Verfilmen sind theoretisch alle Seiten gerade ausgerichtet und in der Größe normiert sowie Helligkeit und Kontraste eingestellt worden. Beim Scannen müssen diese Einstellungen – zumindest in der Theorie – nicht mehr vorgenommen werden, so dass spezielle Einzugsscanner für Mikrofilme verwendet werden können, die mit Geschwindigkeiten von einem Bild pro Sekunde arbeiten.⁹⁷ Dies ermögliche, so das Ziel beispielsweise des TIDEN-Projekts in Skandinavien, einen unkomplizierten und schnellen Workflow. Hier wie unter anderem auch in der British Library wurden daher auch Originale (erneut) mikroverfilmt, um Mikrofilme als Quellmaterial für das Scannen herzustellen.⁹⁸ Schließlich sind beim Scannen vom Mikrofilm vorausgehende konservatorische oder buchbinderische Arbeiten nicht mehr notwendig. Und der beispielsweise in der British Library geäußerten Sorge, beim Scannen vom Original könnte dieses Schaden nehmen, wird ebenfalls begegnet.⁹⁹

In vielen Projekten stellte sich jedoch heraus, dass ältere Mikrofilme für die Digitalisierung nicht verwendbar waren. Dies betraf vor allem solche Mikrofilme, die vor dem Erlass

⁹⁶ Vgl. <http://www.loc.gov/ndnp/content-selection.html>; <http://www.neh.gov/grants/guidelines/ndnp.html>.

⁹⁷ Jefcoate: Burney Collection, S. 186.

⁹⁸ Vgl. Shaw: 10 Billion Words, S. 29; Majlis Bremer-Laamanen: Digitisation for Access to Preserved Documents, in: LIBER Quarterly 13 (2003) 2, S. 137-145, hier S. 140.

⁹⁹ Vgl. King: Digital Historic Newspapers Online, S. 65.

und der Durchsetzung von Standards wie den „IFLA Guidelines for the Preservation Microfilming of Newspapers“¹⁰⁰ oder den „DFG-Vorgaben zur Verfilmung historischer Zeitungsbestände“ angefertigt wurden.¹⁰¹ Zum einen traten Probleme zutage, die auf Fehler bei der Verfilmung schließen ließen, wie beispielsweise eine schlechte Bildqualität oder ein Durchscheinen des Drucks von der Rückseite. Teilweise waren aufgrund von Produktions- oder Entwicklungsfehlern auf den Filmen gar keine Bilder mehr vorhanden. Schwierigkeiten bereiteten auch ungleiche Schärfeverteilung, ungleiche Helligkeit sowie schwarze Flecken auf den Bildern. Zum anderen traten die inzwischen bekannten Nachteile und Qualitätsmängel des bis in die 1970er Jahre hinein verwendeten Zellulose-Acetat-Mikrofilms auf. Dieser ist – anders als der heute verwendete Polyesterfilm – für eine Reihe von Schäden anfällig: Erstens reißt er leichter, zweitens besteht die Gefahr des so genannten „Essig-Syndroms“ (vinegar syndrome) und drittens kann es zum Redox-Phänomen kommen, das farbige Flecken bewirkt. Unsachgemäße Lagerung schließlich führt leicht zu Randwelligkeit und damit zu Unschärfe.¹⁰² Sofern solche fehlerhaften und für die Digitalisierung nicht verwendbaren Mikrofilme vorliegen, bestehen zwei Optionen. Entweder kann eine umfassende erneute Mikroverfilmung erfolgen. Oder man kann, wie beispielsweise beim Projekt „Exilpresse Digital“, aus diesem Grund vom Original scannen.¹⁰³

Für das Scannen vom Papieroriginal werden folgende Argumente verwendet: Primär wird argumentiert, da jeder Medienwechsel einen Qualitätsverlust bedeute, werde eine höhere Qualität der Bilddateien erreicht, wenn nicht vom Mikrofilm, sondern direkt vom Original gescannt werde.¹⁰⁴ Zweitens wird angeführt, beim Fehlen von geeigneten Mikrofilmen sei der Aufwand einer (erneuten) Verfilmung im Vergleich zur „direkten“ Digitalisierung zu groß. Schließlich kann auf Aspekte der Bestandserhaltung abgehoben werden. So wird im Rahmen des österreichischen Projekts ANNO vom Original gescannt –

¹⁰⁰ IFLA Round Table on Newspapers/IFLA Section on Serial Publications (Hrsg.): Guidelines for Newspaper Preservation Microfilming, Den Haag 1996 (IFLA Professional Reports, Bd. 49).

¹⁰¹ Die meisten entsprechenden Normen und Vorgaben wurden erst in den 1980er oder späten 1990er Jahren erlassen. Vgl. auch: Kenning Arlitsch/John Herbert: Microfilm, Paper, and OCR. Issues in Newspaper Digitization at the Utah Digital Newspapers Program, in: Microform & Imaging Review 33 (2004) 2, S. 59-67.

¹⁰² Vgl. Thomas A. Bourke: The Curse of Acetate; or a Base Conundrum Confronted, in: Microform Review 23 (1994) 1, S. 15-17; Helen Shenton: Addressing Cellulose Acetate Microfilm from a British Library Perspective, in: LIBER Quarterly 15 (2005) 2, <http://liber.library.uu.nl/publish/issues/2005-2/index.html?000134>.

¹⁰³ Seib: Exilpresse digital, S. 34; Shaw: 10 Billion Words, S. 29.

¹⁰⁴ Vgl. bspw. Else Delauney: Newspaper Digitisation in France: Present Developements, in: Hartmut Walravens (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 49-53.

und eines der Kriterien für die Aufnahme eines Titels in das Programm ist explizit das Fehlen oder die geringe Qualität von Mikrofilmen.¹⁰⁵

Die zweite wichtige Frage betrifft die eigentlichen Scan-Parameter. Die Entscheidung für eine bestimmte Farbtiefe und Auflösung der Scans ist zum einen von der Überlegung abhängig, wie viel Speicherkapazität jeweils benötigt wird, zum anderen vom intendierten Verwendungszweck. In der Literatur herrschen unterschiedliche Ansichten darüber vor, welche Scan-Parameter die Genauigkeit der OCR-Software begünstigen oder beeinträchtigen. Beim Graustufen-Scan werden erheblich mehr Farbinformationen gespeichert, was, so beispielsweise die Ansicht von Kenning Arlitsch und John Herbert, zu schlechteren OCR-Ergebnissen führen könne.¹⁰⁶

Auf der anderen Seite wurde bei der Erstellung der technischen Vorgaben für das NDNP davon ausgegangen, dass Graustufen-Bilder die bessere Vorlage für die OCR bilden; darüber hinaus sei zu erwarten, dass zukünftige OCR-Software mit Graustufen wesentlich bessere Ergebnisse liefern könne als mit bitonalen Bildern.¹⁰⁷

Hinsichtlich der Farbtiefe sowie der Auflösung, in der gescannt wird, hat sich mittlerweile international ein gewisser Konsens etabliert. Farbdigitalisierung ist für historische Zeitungen, die im Normalfall keine farbigen Elemente enthalten, irrelevant. Stattdessen wird vor allem in einer Farbtiefe von 8 Bit in Graustufen gescannt, der Scan in einer Farbtiefe von 1 Bit in Schwarz-Weiß ist deutlich seltener und das teilweise vorgenommene Scannen in einer Farbtiefe von 4 Bit stellte lediglich eine den technischen Gegebenheiten, insbesondere den Kosten für den Speicherplatz geschuldete Übergangslösung dar.

In Graustufen wird zumeist mit 300 bis 400 dpi gescannt (British Newspaper Pilot, NDNP, digitalnewspapers.org in Utah u. a.), kaum bitonal mit bis zu 600 dpi (DFG-Praxisregeln u. a.). In der British Library wurde der interessante Versuch unternommen, hybride Bild-Dateien zu erstellen, in denen der Textteil bitonal, graphische Elemente hingegen in Graustufen gespeichert sind.¹⁰⁸ Ergebnisse dieser Entwicklung hat die British Library jedoch nicht veröffentlicht.

¹⁰⁵ Vgl. Christa Müller: ANNO – AustriaN Newspapers Online. Historische österreichische Zeitungen und Zeitschriften online. Eine Digitalisierungsinitiative der Österreichischen Nationalbibliothek, in: Hartmut Walravens (Hrsg.): Newspapers in Central and Eastern Europe. Zeitungen in Mittel- und Osteuropa. Papers Presented at an IFLA Conference held in Berlin, August 2003, München 2005, S. 141-148.

¹⁰⁶ Vgl. Arlitsch/Herbert: Microfilm. Auch Ralf Stockmann, Leiter des Göttinger Digitalisierungszentrums bestätigt, dass für die OCR bitonales Scannen unter Umständen zu den besseren Ergebnissen führe.

¹⁰⁷ Ray L. Murray: The U.S. National Digital Newspaper Program: Thinking Ahead, Designing Now, Part III: Technical Specifications, in: Hartmut Walravens (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 87-89.

¹⁰⁸ Vgl. King: Digital Historic Newspapers Online, S. 66.

Hinsichtlich der (Daten-)Formate der beim Scan erzeugten Dateien hat sich international ein recht einheitlicher Standard etabliert. Grundsätzlich wird zwischen Master-Dateien für die Archivierung (und ggf. Produktion) und Derivaten für die Präsentation unterschieden. Als Master-Dateien werden in der Regel Tiff-Dateien¹⁰⁹ (unkomprimiert oder mit Group4-Komprimierung) erzeugt, da diese gegenüber anderen Formaten eine Reihe von Vorteilen besitzen: Sie lassen sich unter anderem verlustfrei bearbeiten und speichern, sind nicht proprietär, weit verbreitet (beziehungsweise implementiert) und lassen die Speicherung von Metadaten im Datei-Header zu. Nach heutigem Erkenntnisstand sind sie so auch für eine digitale Langzeitarchivierung geeignet.¹¹⁰ Eine Alternative scheint inzwischen das Format Jpeg2000 darzustellen. Daher wurde in Folge einer Evaluierung der bisherigen Digitalisierungsaktivitäten an der Königlichen Bibliothek der Niederlande eine derzeit noch laufende Studie in Auftrag gegeben, die die Verwendung von Jpeg2000 als Alternative zu Tiff untersuchen soll.¹¹¹ Eine entsprechende Untersuchung liegt auch für das NDNP vor, in dessen Folge die technischen Vorgaben der Library of Congress für die NDNP-Projektteilnehmer die Verwendung von Jpeg2000 als „production master“ (im Unterschied zum „archival master“) vorsehen.¹¹²

b. Erschließung

Hochwertige Bilddateien allein sind, in der Diktion von Gerd-Josef Bötte, langfristig lediglich „Elektroschrott“.¹¹³ Für eine sinnvolle und die Erstellung rechtfertigende Nutzung müssen sie in ausreichender Tiefe erschlossen werden. Bei dieser Erschließung von Zeitungen lassen sich hauptsächlich die Gesichtspunkte Volltexterfassung, Artikelseparierung und Strukturdatenerfassung sowie die verwendeten Metadatenformate unterscheiden. Diese Aspekte sollen im Folgenden betrachtet werden.

Einen wesentlichen Mehrwert digitalisierter Textdokumente gegenüber analogen Trägern bietet die Möglichkeit, in einem definierten Korpus nach einer bestimmten Zeichenfolge zu suchen. Hierzu muss der Textinhalt aus der Bilddatei in eine maschinenlesbare Form gebracht werden. Während dies bis vor einigen Jahren mit zufrieden stellenden

¹⁰⁹ Vgl. <http://partners.adobe.com/public/developer/tiff/index.html>.

¹¹⁰ Vgl. Uwe Borghoff u. a.: Langzeitarchivierung. Methoden zur Erhaltung digitaler Dokumente, Heidelberg 2003, S. 110-112; nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, hrsg. von Heike Neuroth u. a., <http://nestor.sub.uni-goettingen.de/handbuch/index.php>, Kapitel 15.2: Bilddokumente.

¹¹¹ Vgl. Verheusen: Newspaper Digitisation in the Netherlands, S. 59-60.

¹¹² Vgl. JPEG 2000 Profile for the National Digital Newspaper Program, bearbeitet von Robert Buckley/Roger Sam, Library of Congress/Office of Strategic Initiatives, 27. April 2006.

¹¹³ Gerd-Josef Bötte: Das deutsche Schrifttum des achtzehnten Jahrhunderts und seine bibliographische Verzeichnung – Perspektiven für ein VD18, in: Daniela Lülfig/Irmgard Siebert (Hrsg.): „Geld ist rund und rollt weg, aber Bildung bleibt“. 94. Deutscher Bibliothekartag in Düsseldorf 2005, Frankfurt am Main 2006 (ZfBB Sonderheft 89), S. 155-162.

Ergebnissen lediglich manuell durch Abtippen der Texte möglich war, werden ausreichend hohe Erkennungsgenauigkeiten inzwischen auch durch eine automatische Texterkennung erreicht.

Für den Begriff der automatischen Texterkennung wird häufig der Terminus „Optical Character Recognition“ (OCR) synonym verwendet. Streng genommen bildet die OCR nur einen Teil der Texterkennung, nämlich die eigentliche Mustererkennung. Dabei werden pixelweise Muster gesucht und mit in der Software gespeicherten Mustern abgeglichen. Diese sind einzelnen Zeichen zugeordnet, so dass anhand erkannter Muster ein Text (als Abfolge von Zeichen) ausgegeben werden kann. Texterkennung umfasst darüber hinaus auch „Intelligent Character Recognition“ (ICR), in der die Ergebnisse der OCR durch eine Kontextanalyse korrigiert werden.

Als Mindestgenauigkeit, die bei der Texterkennung zu erzielen sei, finden sich in der Literatur verschiedene Angaben, die von 95% bis 99,5% reichen, wobei zumeist unklar bleibt, ob es sich um die Genauigkeit der Erfassung von Zeichen oder Wörtern handelt.¹¹⁴ Schwierigkeiten speziell bei Zeitungen bereitet insbesondere die – im Vergleich zu monographischen Schriften (und Zeitschriften) – ungleich komplexere Seitenstruktur. Zeitungen des 17. und 18. Jahrhunderts sind vielfach in ein bis zwei Spalten mit fortlaufendem Text gesetzt. Logische Einheiten sind im Idealfall durch Absätze oder kurze Trennstriche gekennzeichnet, nicht aber durch Überschriften. Im Laufe des 19. Jahrhunderts wurden Zeitungen größer und enthielten mehr Spalten, in denen inhaltliche Abschnitte unter Umständen deutlicher getrennt waren; diese verliefen aber nicht mehr notwendigerweise linear von links oben nach rechts unten. Darüber hinaus erscheinen hier mit Abbildungen (und deren Beschriftung) sowie Werbung neue Elemente. Schließlich stellt die Verwendung unterschiedlicher Schriftgrößen, ungewöhnlicher Schrifttypen sowie verschiedener Schriftauszeichnungen für jede OCR-Software Probleme dar.¹¹⁵

Weitere Schwierigkeiten erwachsen aus dem Druck mit zu wenig oder zu viel Tinte, aus schief gedruckten Seiten, von der Rückseite durchscheinenden Zeichen, Verschmutzungen und Beschädigungen. Dies sind keine auf Zeitungen beschränkten Phänomene, sie treten hier gleichwohl besonders stark und häufig auf – aufgrund des dünnen und schlechten Papiers sowie des schnellen und nicht auf hohe Qualität ausgerichteten Drucks.

¹¹⁴ Eine Genauigkeit von 95% der Zeichen bedeutet, dass jedes zwanzigste Zeichen falsch ist. Die Fehlerquote der Wörter ist dabei in der Regel höher. Geht man davon aus, dass die durchschnittliche Wortlänge im Deutschen sechs Zeichen beträgt, bedeutet eine Genauigkeit von 95% der Zeichen, dass nur 73% der Wörter korrekt sind. Die Zeichengenauigkeit ist somit deutlich höher, für die Recherche ist jedoch ausschließlich eine ausreichend hohe Wortgenauigkeit erforderlich. Vgl. Simon Tanner: Measuring the OCR Accuracy Across The British Library's 2 Million Page Newspaper Archive, Vortrag auf der IMPACT-Konferenz, April 2009. 1997 stellten Marianne Dörr und Hartmut Weber fest, erst ab einer Genauigkeit von 99,5% sei eine automatisierte Texterfassung günstiger als eine manuelle, also bei weniger als fünf Fehlern auf 1.000 Zeichen. Vgl. Marianne Dörr/Hartmut Weber: Digitalisierung als Mittel der Bestandserhaltung? Abschlußbericht einer Arbeitsgruppe der Deutschen Forschungsgemeinschaft, in: Zeitschrift für Bibliothekswesen und Bibliographie 44 (1997) 1, S. 53-76.

¹¹⁵ Vgl. bspw. Shaw: 10 Billion Words, S. 33-34.

Dieser resultierte unter Umständen auch in schattierten Hintergründen, fragmentierten, sich berührenden oder überlappenden Buchstaben. Und schließlich führt die häufig zu beobachtende enge Bindung von Zeitungsbänden dazu, dass nach dem Scannen Zeilen schief oder gekrümmt erscheinen.¹¹⁶

Im deutschen Sprachraum stellt sich die Situation noch einmal als besonders kompliziert dar. Während in anderen Staaten Zeitungen in Antiqua-Schriften gedruckt wurden, herrschte in Deutschland bis 1941 die Verwendung von Frakturschriften vor. Aufgrund des Aufbaus der Buchstaben, der Verwendung von teilweise sehr feinen Linien sowie der Ähnlichkeit verschiedener Buchstaben liegt hier die Genauigkeit automatischer Erfassung um ein Vielfaches niedriger als bei Antiqua-Schriften.¹¹⁷

In bereits existierenden Projekten wurde die Problematik ungenauer Ergebnisse der Volltextfassung unterschiedlich gelöst. Die komplette manuelle Nachbearbeitung wird vielfach als zu aufwändig und teuer erachtet und daher verworfen.¹¹⁸ Sie wird vor allem in den privatwirtschaftlichen Projekten vorgenommen. Ausgehend von der Relevanz einzelner Textabschnitte (für das Retrieval) ist aber auch ein manuelles „re-keying“ von Titel und Untertitel jedes Artikels denkbar. Das australische Zeitungsdigitalisierungsprojekt ANDP setzt ebenfalls auf eine manuelle Korrektur des automatisch erstellten Textes, verfolgt dabei allerdings einen kollaborativen Ansatz. Nutzern ist es möglich, sich auf dem Portal des Projekts zu registrieren und in einer intuitiv zu bedienenden Arbeitsumgebung selbständig, anhand des Images, Volltexte der Artikel zu korrigieren.¹¹⁹ Ein anderer Ansatz schließlich, der weniger auf die Erstellung des Textes als auf die „Suchbarkeit“ abhebt, ist die Verwendung von „fuzzy logic“, bei der die Ungenauigkeit der Volltextfassung durch eine Unschärfe bzw. Toleranz der Suche zu kompensieren versucht wird.¹²⁰

¹¹⁶ Arlitsch/Herbert: Microfilm, S. 63; Verheusen: Newspaper Digitisation, S. 58-60.

¹¹⁷ Für einen Vergleich und eine Bewertung von OCR-Software sei an dieser Stelle auf die einschlägige Literatur verwiesen. Wesentlich gerade für den deutschen Sprachraum ist darüber hinaus die Tatsache, dass der Markt an OCR-Software, die instande ist, Frakturtexte zu erfassen, sehr klein ist. Letztlich sind hier vier Produkte zu nennen: OCRopus (Google), FineReader XIX (Abbyy), BIT-Fraktur (Bureau Ingenieur Tomasi) sowie die Software der Firma PPS. In einer vergleichenden Untersuchung der drei erstgenannten Produkte kommt Thomas Stäcker zu dem Ergebnis, dass derzeit lediglich BIT-Fraktur – mit Training und Nachbearbeitung – Genauigkeiten von 99% zu erreichen instande ist. Vgl. Thomas Stäcker: OCR für Alte Drucke – eitle Hoffnung oder konkrete Perspektive? Vortrag, Blaubeuren, 21.10.2008, <http://www.bibliothekerverband.de/aghandschriften/dokumente/2008-staecker.pdf>.

¹¹⁸ Als durchschnittliche Kosten für ein manuelles (re-)keying wird pro Seite ungefähr ein Euro veranschlagt. Vgl. <http://www.impact-project.eu/about-the-project/concept/>.

¹¹⁹ Vgl. Rose Holley: How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs, in: D-Lib Magazine 15 (2009) 3/4, <http://www.dlib.org/dlib/march09/holley/03holley.html>; Rose Holley: The Australian Newspapers Digitisation Program. Helping Communities Access and Explore their Newspaper Heritage, Vortrag auf der Australian Media Traditions Conference, Bathurst 23.11.2007, http://www.nla.gov.au/nla/staffpaper/2008/documents/ANDP_AMTKeynotePaper_Nov2007.pdf.

¹²⁰ Vgl. King: 19th Century British Library Newspapers, S. 53. Für andere Verfahren zur Verbesserung der Suchergebnisse, über deren Verwendung in aktuellen Projekten der Zeitungsdigitalisierung nichts

Anfang 2008 wurde durch die EU das Projekt „Improving Access to Text“ (IMPACT) initiiert. Unter Federführung der Königlichen Bibliothek der Niederlande forschen europäische Bibliotheken, Unternehmen und Forschungseinrichtungen daran, Methoden und Techniken der Textextraktion zu verbessern. Gerade für die Volltexterkennung von Frakturschriften wird von diesem Projekt eine wegweisende Entwicklung erwartet.¹²¹

Volltexte werden vielfach in den Formaten TEI (der Text Encoding Initiative) oder ALTO (Analyzed Layout and Text Object; entwickelt aus dem METAe-Projekt) gespeichert. Insbesondere im deutschsprachigen Raum hat sich inzwischen vielfach ALTO als Standardformat für Volltexte etabliert. ALTO ermöglicht es, nicht nur die Texte selbst zu speichern, sondern auch einzelne Zeichen mit Koordinaten der Bilddatei zu verknüpfen. Dadurch können unter anderem Wörter in der Bilddatei markiert beziehungsweise graphisch hervorgehoben werden, beispielsweise als Ergebnis einer Suchanfrage. Gerade für großformatige Zeitungen ist dies ein nützliches Instrument.

Ein weiterer wichtiger Schritt des Digitalisierungs-Workflows ist die strukturelle inhaltliche Erschließung einzelner Elemente. Grundlegend ist dabei die Erfassung der – in der Regel bereits in einem maschinenlesbaren Format vorliegenden – bibliographischen Daten der Zeitung. Die Erfassung kann dann weiter heruntergebrochen werden und es können einzelne Ausgaben, deren Teile oder Artikel einzeln erschlossen werden. Gerade hinsichtlich dieser Erschließung unterscheiden sich die verschiedenen Projekte in der Tiefe und in dem Verfahren der Erschließung. Im internationalen Vergleich zeigt sich, dass die Separierung einzelner Inhaltselemente, vor allem der Artikel, aber auch von Anzeigen, Bildern oder Bildunterschriften, zum Standard geworden ist.¹²² Diese Elemente werden einzeln erschlossen und mit Angaben wie insbesondere Datum und der Überschrift versehen. Die Erfassung von Überschriften kann allerdings Schwierigkeiten bereiten, da viele Zeitungsartikel keine solche besitzen. Eine Lösung ist in diesem Fall, aus dem Anfang des Artikels eine Überschrift zu fingieren. Auch Todes- oder Geburtenanzeigen, Annoncen, Werbung und Ähnliches lassen sich schwierig unter eine Überschrift fassen. Problematisch ist schließlich auch die Tatsache, dass Zeitungsartikel unter Umständen auf einer Seite beginnen und auf einer oder mehreren, auch nicht zusammenhängenden Seiten fortgesetzt werden.

Die Artikelseparierung erfolgt (halb)automatisch; die Software erkennt die Struktur einer Seite und setzt sie in logische Einheiten um. An Techniken und Methoden zur automatischen Separierung und Extraktion von Zeitungsartikeln wird in vielen Ländern geforscht. Da diese Ansätze vielfach auf ähnlicher Technologie basieren, waren sie

verlautbart wird, vgl. Thomas Parschik: Durchführung von Digitalisierungsprojekten in Bibliotheken, in: Bibliotheksdienst 40 (2006) 12, S. 1421-1444, insbes. S. 1435-1440.

¹²¹ Vgl. <http://www.impact-project.eu/>.

¹²² Vgl. bspw. King: 19th Century British Library Newspapers.

gleichermaßen davon betroffen, dass die niederländische Firma Océ-Technologies einen „Apparatus and method for extracting articles from a document“ zum Patent anmeldete. Der Patentschutz besteht seit 1996 in Deutschland, Frankreich, Großbritannien, Italien und den Niederlanden und seit 2000 in den USA.¹²³ Noch kann nicht abgesehen werden, wie sich diese Patente auf die weitere Entwicklung von Software zur Artikelseparierung auswirken. Zumindest dürften sie die Weiterentwicklung hemmen.

Auch für die Erfassung von Metadaten hat sich ein Set an Formaten und Standards herausgebildet, wenngleich hier eine größere Diversität herrscht. Michael Seadle hat in diesem Zusammenhang vom „metadata marketplace“ einer Vielzahl konkurrierender Formate gesprochen.¹²⁴ Im Rahmen dieser Arbeit liegt der Schwerpunkt auf solchen Formaten, die sich in bisherigen Zeitungsdigitalisierungsprojekten als Standard herauskristallisiert haben.¹²⁵

Dies trifft in besonderem Maße auf METS (Metadata Encoding and Transmission Standard) zu.¹²⁶ METS erlaubt es, unter anderem technische, bibliographische, deskriptive und strukturelle Metadaten in einer Datei abzulegen. Die „deskriptiven“ Daten werden dabei allerdings nicht in einem METS-eigenen Form abgelegt, sondern in anderen, hierfür genuin vorgesehenen Formaten – beispielsweise MODS¹²⁷ (Metadata Object Description Schema).¹²⁸

Für Periodika und insbesondere Zeitungen ist die Möglichkeit interessant, Strukturen nicht nur innerhalb einer einzelnen METS-Datei abzubilden. Bei hierarchischen, logischen Strukturen kann für jedes Element oder Objekt eine eigene METS-Datei angelegt werden, die miteinander verknüpft werden können. So können Metadaten im METS-Format für jeden Artikel einer Zeitung abgelegt werden, zusätzlich für jede Ausgabe und schließlich für die gesamte Zeitung.

Ein anderer interessanter Aspekt von METS ist die Option, spezielle, auf die Bedürfnisse einzelner Projekte oder Institutionen ausgerichtete METS-Profile zu erstellen und bei der Library of Congress, der Koordinationsstelle des METS-Standards, zu registrieren. Bisher sind dort allerdings keine explizit auf Zeitungen ausgerichteten Profile vorhanden.¹²⁹ Die Erstellung eines auf (digitalisierte historische) Zeitungen ausgerichteten

¹²³ Europäisches Patent EP 0753833 B1, erteilt am 26.6.1996; U.S. Patent 6,134,565, erteilt am 17.10.2000.

¹²⁴ Michael Seadle: METS and the Metadata Marketplace, in: Library Hi-Tech 20 (2002) 3, S. 255-257.

¹²⁵ Andere Formate, wie vor allem Dublin Core, werden derzeit für digitalisierte Zeitungen nur in wenigen Fällen verwendet, sie werden daher hier in die Betrachtung nicht mit einbezogen.

¹²⁶ Vgl. <http://www.loc.gov/standards/mets/>; Ray L. Murray: Toward a Metadata Standard for Digitized Historical Newspapers, <http://www.loc.gov/ndnp/MetadataStandard.html>.

¹²⁷ Vgl. <http://www.loc.gov/standards/mods/>.

¹²⁸ Vgl. Morgan Cundiff: An Introduction to the Metadata Encoding and Transmission Standard (METS), in: Library Hi Tech 22 (2004) 1, S. 52-64.

¹²⁹ Vgl. <http://www.loc.gov/standards/mets/mets-profiles.html>; für einen Entwurf eines Profils für historische Zeitungen vgl. Morgan Cundiff: A METS Application Profile for Historical Newspapers, in:

Profils sowie dessen Registrierung könnten richtungsweisend für die Etablierung internationaler Standards und darauf aufbauender Rechercheoptionen sein.

c. Präsentation und Archivierung

Der Prozess der Zeitungsdigitalisierung verfolgt, wie in der Einleitung dargelegt, im Wesentlichen zwei Ziele: Die digitalisierten Dokumente sollen zum einen einem möglichst breiten Kreis an Nutzern zugänglich gemacht, zum anderen dauerhaft gesichert werden. Präsentation sowie Speicherung und Archivierung sind daher zwei wichtige Fragen im Rahmen des Digitalisierungsprozesses und sollen im Folgenden erörtert werden. Möglichkeiten der Präsentation und der Suchoptionen für digitalisierte Zeitungen wurden bereits im Rahmen des Vergleichs der Digitalisierungsstrukturen in den Niederlanden, Großbritannien und den USA kurz skizziert. Diese Aspekte sollen hier zusammengefasst und darauf aufbauend übergeordnete Aussagen entwickelt werden.

Die meisten Projekte gestalten die Software zur Präsentation der Zeitungen nach zwei Prinzipien, die von der British Library explizit formuliert wurden: Die Digitalisate sollen „readable“ und „searchable“ sein. Das heißt, Benutzer sollen in den digitalisierten Zeitungen wie in der Printausgabe blättern können und den gleichen Gesamteindruck von jeder einzelnen Seite erhalten wie im Original. Zugleich sollen jedoch die Möglichkeiten genutzt werden, die sich aus der Erschließung der Zeitungen, der Strukturdatenerfassung, aus der gegebenenfalls vorgenommenen Artikelseparierung oder Volltexterfassung ergeben. Als Standard hat sich der Rechercheeinstieg nach Zeitung, Datum und/oder Ausgabe herausgebildet. In den nicht deutschsprachigen Gebieten gehört darüber hinaus die Erschließung auf der Ebene einzelner Artikel zum Standard, wobei zumeist Datum, Überschrift und ggf. Verfasser angegeben werden. Einige Projekte gehen in der Tiefe der Erschließung hierüber noch hinaus und weisen einzelne Inhaltselemente zuvor definierten Kategorien zu, wie beispielsweise Nachricht, Illustration, Todesanzeige, Werbung oder anderen. Schließlich kann durch den Zugriff auf (jeweils artikelweise) gespeicherte Volltexte eine wesentlich genauere und zielgerichtete Suche ermöglicht werden. Für die Präsentation werden in der Regel Jpeg-Dateien verwendet, für Druck und Download auch PDF-Dateien, teils bestehend aus einem Bild und dahinterliegendem Text.

Hinsichtlich der Speicherung und Archivierung der digitalisierten Zeitungen finden sich generell relativ wenige Angaben zu den einzelnen Projekten. Die veröffentlichten Informationen betreffen vor allem die Frage, wo die Digitalisate gespeichert werden – in den Gedächtniseinrichtungen oder in Rechenzentren, in der digitalisierenden Institution oder an

Hartmut Walravens (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 135-140.

zentraler Stelle. Die Frage nach einer zentralen Speicherung stellt sich besonders den Staaten, die keine zentral digitalisierende Instanz aufweisen. In den USA beispielsweise werden die in den einzelnen Bundesstaaten im Rahmen des NDNF generierten Digitalisate an die Library of Congress überspielt und dort gespeichert.

Welche Konzepte der digitalen Langzeitarchivierung (LZA) verfolgt werden, ist bisher noch weniger bekannt. Fast alle betrachteten Projekte geben an, sich bei der Gestaltung des Workflows und der (Meta-)Datenstruktur an den Vorgaben des „Open Archival Information System“ (OAIS) zu orientieren. Dies ist derzeit das internationale Referenzmodell für die LZA. Die Königliche Bibliothek in Den Haag hat über mehrere Jahre das so genannte eDepot aufgebaut, das der Archivierung digitaler Daten dient und in das die digitalisierten Zeitungen eingespielt werden sollen. Für andere Staaten sind vergleichbare Planungen nicht veröffentlicht.

3. Zusammenfassung und Einordnung

In der vorliegenden Untersuchung wurden Rahmenbedingungen und Projekte der Zeitungsdigitalisierung in ausgewählten Staaten verglichen und anschließend der Stand der Technik anhand kritischer Teilaspekte betrachtet. Dabei konnten auf internationaler Ebene interessante Unterschiede und Gemeinsamkeiten herausgearbeitet werden.

Die größten Unterschiede ergeben sich aus den nationalspezifischen Rahmenbedingungen, da die geschaffenen Projektstrukturen hierauf aufbauen. Grundsätzlich lassen sich folgende Modelle beobachten: Das erste Modell findet sich unter anderem in Großbritannien und beinhaltet die zentrale Abwicklung des Programms über eine Bibliothek. Der komplette Prozess der Planung und Durchführung wird dann von einer zentralen Stelle gesteuert, andere Gedächtnisinstitutionen werden gegebenenfalls als Lieferanten benötigter Bestände involviert. Weitere Beispiele für dieses Modell finden sich in Österreich und in Australien. Modell 2 beinhaltet zwar eine zentrale Steuerung, die jedoch über die Finanzierung und die daran geknüpften (technischen) Vorgaben erfolgt. Dies findet sich in den USA, entspricht aber vom Prinzip her weitgehend auch dem Ansatz, den die DFG – zwar nicht speziell hinsichtlich der Zeitungsdigitalisierung, aber beispielsweise auf dem Gebiet der Mikroverfilmung – verfolgt. Drittens schließlich (Modell 3) besteht aktuell in einigen Ländern die Situation, dass sich bereits eine Vielzahl von Digitalisierungsprojekten entwickelt hat und eine zentrale Instanz nun versucht, über eigene groß angelegte Projekte sowie die Einrichtung einer landesweiten Initiative steuernd und vereinheitlichend zu wirken.

Unterhalb dieser grundsätzlichen organisatorischen Ebene lassen sich mehr Gemeinsamkeiten als Unterschiede ausmachen. Erstens herrscht Einigkeit darüber, dass nicht der komplette Zeitungsbestand eines Landes digitalisiert werden könne und daher eine repräsentative Auswahl getroffen werden müsse. Hierzu wurden in den Vergleichsstaaten einerseits beratende Gremien gebildet, andererseits klare Kriterienkataloge erarbeitet. Diese sehen durchaus unterschiedlich aus, dennoch weisen sie mehrere allen gemeinsame Aspekte auf: Es wurden sowohl nationale als auch lokale und regionale Zeitungen ausgewählt und es sollte die gesamte „Lebenswirklichkeit“ der jeweiligen Zeit und des jeweiligen Landes berücksichtigt werden, sofern sie sich in Zeitungspublikationen niederschlug – das heißt politische oder geistesgeschichtliche Strömungen, Bevölkerungsgruppen und -zentren. Zweitens wird in allen aktuellen Projekten ein großer Teil, wenn nicht die gesamte Digitalisierung an Dienstleister vergeben. In der eigenen Institution werden lediglich die Projektleitung und das Qualitätsmanagement sowie gegebenenfalls die Arbeitsschritte Präsentation und Archivierung angesiedelt.

Überwiegend Gemeinsamkeiten zeigen sich auch hinsichtlich des Stands der Technik – hier haben sich vielfach Standards oder „standardisierte“ Entscheidungsalternativen herausgebildet: Ob vom Mikrofilm oder vom Papieroriginal gescannt wird, hängt von einer Reihe von Faktoren ab, so dass die Entscheidung immer im Einzelfall zu treffen sein wird. International wird bevorzugt vom Mikrofilm gescannt, es finden sich jedoch überzeugende Argumente für beide Varianten. In Bezug auf Scan-Parameter, Daten- und Metadatenformate kann ein weitgehender Konsens festgestellt werden. Auch hinsichtlich der Erschließung lassen sich internationale Standards erkennen. Dies ist zum einen die Separierung einzelner Artikel sowie deren individuelle Erschließung, zum anderen die Volltexterfassung.

Die vorliegende Untersuchung hat gezeigt, dass sich in vielen organisatorischen und technischen Fragen entweder bereits Standards für die Zeitungsdigitalisierung herausgebildet haben oder klare Entscheidungsalternativen benannt werden können. Hierauf kann bei der Konzeption zukünftiger Projekte aufgebaut werden, wobei diese Standards sowie die Alternativen für jeden Aspekt des Digitalisierungsprozesses einzeln überprüft werden müssen: Sind sie angesichts der konkret vorhandenen Bedingungen anzupassen, zu verwerfen oder zu übernehmen?

C. Die Perspektive: Ein Zeitungsdigitalisierungsprogramm für Deutschland

1. Grundlagen und Vorbereitungen

Auf der Grundlage des in dieser Arbeit vorgenommenen internationalen Vergleichs können nun fundierte Vorschläge für ein nationales Zeitungsdigitalisierungsprogramm in Deutschland (im Folgenden ZDPD) entwickelt werden. Die Struktur dieses Programms muss dabei auf die spezifischen deutschen Verhältnisse zugeschnitten sein. Für ein solches Programm bestehen mehrere Sachzwänge: Zum einen ist die Grundvoraussetzung für die Realisierbarkeit ein belastbares Finanzierungskonzept, bei dem die Last der Finanzierung idealiter auf die Schultern mehrerer Institutionen und über einen längeren Zeitraum verteilt wird. Dies erhöht die Durchsetzbarkeit und Sicherheit des Programms, da so eine stetige Finanzierung eher gesichert ist. Zum anderen wurde im Vergleich der Digitalisierungsprojekte anderer Staaten deutlich, dass nicht der gesamte Zeitungsbestand eines Landes mit einem Male digitalisiert werden kann und daher eine Auswahl zu treffen ist.

In den Projekten anderer Staaten hat sich gezeigt, dass die Einteilung eines groß angelegten Zeitungsdigitalisierungsprogramms in mehrere Phasen diesen Umständen Rechnung trägt, da hierdurch eine finanzielle und organisatorische Entzerrung erzielt wird. Auch für Deutschland soll daher ein Programm entwickelt werden, das in mehreren Phasen abläuft.

Generell ist die Durchführung, sogar allein die Vorbereitung eines solchen Programms nicht denkbar ohne die Kenntnis der vorhandenen Bestände. Denn nur auf der Grundlage einer systematischen Erfassung der Zeitungsbestände kann eine begründete Auswahl von Zeitungen für die Digitalisierung vorgenommen werden. Zwar sind Bestandsverzeichnisse in vielen Bibliotheken vorhanden, diese sind jedoch bei weitem nicht vollständig. Darüber hinaus wird ein Überblick über die in Deutschland publizierten Zeitungen sowie über die noch vorhandenen Bestände benötigt, um beurteilen zu können, inwiefern eine Zeitung beziehungsweise ein Zeitungsbestand repräsentativ (für eine Gruppe von Zeitungen, für eine Region u. a.) ist. Wünschenswert ist daher die Durchführung des seit vielen Jahren wiederholt von Zeitungsbibliothekaren und Presseforschern angemahnten Programms zur Bestandserhebung und -erschließung. Da ein solches Programm jedoch, wie die Beispiele der entsprechenden Programme in anderen Ländern gezeigt haben, zeitaufwändig und kostenintensiv ist, sollte es nicht zu einer *conditio sine qua non* für das geplante ZDPD

gemacht werden. Denn dies könnte das ZDPD erheblich verzögern oder insgesamt in Frage stellen. Schließlich hat das Vorgehen der Königlichen Bibliothek der Niederlande gezeigt, dass auch ohne ein solches Programm breit angelegte Digitalisierungsvorhaben durchgeführt werden können – wenngleich auf einer dünneren Faktenbasis und einer tendenziell weniger repräsentativen Auswahl beruhend.

Für das im Folgenden vorgestellte Digitalisierungsprogramm bietet sich eine Aufteilung in drei Phasen an, in denen zum einen durch zentral gesteuerte Digitalisierungsmaßnahmen eine breite Materialbasis gewährleistet wird, zum anderen aber auch lokale Bestände durch die Förderung dezentraler Projekte Berücksichtigung finden. In der ersten Phase der Digitalisierung soll aus zentralem Bestand ein Kern-Korpus digitalisierter historischer Zeitungen entstehen, in der zweiten Phase erfolgt eine konzertierte Digitalisierung in den Bundesländern und die dritte Phase besteht in der Realisierung einer langfristig angelegten DFG-Aktionslinie „Zeitungsdigitalisierung“, die die Förderung von Einzelprojekten ermöglicht.

2. Erste Phase: Zentrale Digitalisierung

In der ersten Phase des ZDPD erfolgt eine zentrale Digitalisierung eines ersten Korpus historischer Zeitungen. Der Ablauf dieser Phase wird im Folgenden entwickelt, wobei sich die Ausführungen an der Abfolge der Arbeitsschritte orientieren, das heißt: die Konzeption des Vorgehens, die Planung der Finanzierung, Auswahl des zu digitalisierenden Materials, die Erschließung sowie schließlich die technischen Aspekte.

Ziel der ersten Phase ist die zentrale und kostengünstige Digitalisierung eines ersten Korpus' historischer Zeitungen. Hierfür bietet sich die Digitalisierung eines umfangreichen, geschlossenen Bestands an, der in einheitlicher Form vorliegt. Ein solcher Bestand ist, wie in Kapitel B.1.a ausgeführt wurde, im „Mikrofilmarchiv der Deutschsprachigen Presse“ vorhanden. Daher wird vorgeschlagen, die dortigen mikroverfilmten Zeitungen in größerem Umfang zu digitalisieren. Dieses Vorgehen brächte mehrere Vorteile: Erstens kann vom Mikrofilm – wie sich im Rahmen bisheriger Projekten gezeigt hat – schneller und günstiger digitalisiert werden als vom Original. Dadurch könnte eine Verringerung des finanziellen Aufwands gewährleistet und ein zügiger Programmstart mit raschen „Anfangserfolgen“ ermöglicht werden. Zweitens könnte so ein Korpus aufgebaut werden, an dem Methoden der Erschließung zunächst getestet werden können. Drittens kann für dieses Korpus die technische Infrastruktur (Speicherungs- und Präsentationssystem) leichter implementiert werden als für Digitalisate, die von verschiedenen Institutionen

erstellt und an zentraler Stelle zusammengeführt werden müssen. Viertens sind die im MFA vorhandenen Zeitungen in ihren Erscheinungsverläufen vollständig vorhanden, da diese Vollständigkeit eine Voraussetzung für die DFG-Förderung der Mikroverfilmung darstellte. Und fünftens sprechen inhaltliche Gründe hierfür: Das Korpus eignet sich dazu, eine überregionale Perspektive zu erfassen; zugleich kann durch eine gezielte Auswahl sichergestellt werden, dass Regionen berücksichtigt werden, die heute außerhalb des deutschen Staatsgebiets liegen beziehungsweise für die keine Landesbibliothek zuständig ist.

Eine vorbereitende Untersuchung muss jedoch durchgeführt werden: Da ältere Mikrofilme, wie in Kapitel B.2.a dargelegt wurde, aufgrund verschiedener Phänomene unter Umständen nicht mehr für eine Digitalisierung verwendbar sind, muss zunächst der Zustand der Mikrofilme im MFA überprüft werden. Sollte sich dabei ergeben, dass ein überwiegender Teil des MFA-Bestands für die Digitalisierung unbrauchbar ist, wäre zu überlegen, ob nicht andere umfangreiche Mikrofilmbestände, wie beispielsweise in der Staatsbibliothek zu Berlin, alternativ genutzt werden könnten.

Nachdem somit ein Korpus an Zeitungen für die Digitalisierung benannt ist, muss im nächsten Schritt die Finanzierung des Vorhabens sichergestellt werden. Im internationalen Vergleich zeigt sich, dass die Programme in anderen Ländern – beispielsweise in den Niederlanden, Großbritannien und den USA – durch die jeweilige Institution zur Wissenschaftsförderung finanziert werden. Auch für Deutschland ist eine primär bei der DFG angesiedelte Finanzierung sinnvoll, schließlich ist die DFG die zuständige Einrichtung für die Förderung wissenschaftlicher Infrastruktur und Programme und mittlerweile auch für die Bewahrung des kulturellen Erbes. Sie verfügt zudem über die notwendigen Fördermittel und hat bereits eine große Zahl strukturell vergleichbarer Großvorhaben gefördert. Allerdings wird die DFG nicht die komplette Finanzierung aller Programmphasen übernehmen können. Dies wird deutlich, wenn man die Budgets der Programme anderer Länder betrachtet. Die Zeitungsdigitalisierungsprojekte der British Library sind mit zusammen mindestens drei Millionen Britischen Pfund zu veranschlagen, die Königliche Bibliothek der Niederlande erhielt für die „Databank Digitale Dagbladen“ rund 12,5 Millionen Euro und für das NDNF werden Mittel in Höhe von 20 bis 25 Millionen US-Dollar aufgewendet. Im Vergleich hierzu wurden 2007 von der DFG im Rahmen der Förderung der „Wissenschaftlichen Literatur- und Informationsversorgung“ für die gesamte „Kulturelle Überlieferung“ rund 12,6 Millionen Euro bereitgestellt.¹³⁰

Diese Angaben – die auch Anhaltungspunkte für den gesamten Finanzierungsbedarf bieten – belegen die Notwendigkeit einer institutionellen und zeitlichen Entzerrung. Die

¹³⁰ Vgl. Deutsche Forschungsgemeinschaft: Jahresbericht 2007. Aufgaben und Ergebnisse, Bonn 2008, S. 82.

Finanzierung sollte daher je nach Programmphase unterschiedlich ausfallen. Die erste Phase müsste durch Mittel der DFG finanziert werden, da an dieser Phase keine Institutionen beteiligt sind, die zur Finanzierung beitragen könnten. Auch die übliche Regelung der DFG-Förderung, die umfangreiche Eigenleistungen vorsieht, ist hier nicht anwendbar, da das MFA als Projektträger nicht über entsprechende Mittel verfügt. Als Ergänzung zu DFG-Mitteln sollte jedoch die Möglichkeit einer Förderung durch das Bundesministerium für Bildung und Forschung geprüft werden. Diese wäre beispielsweise aus Fördermitteln für die Entwicklung und Implementierung der Grid-Technologie möglich.¹³¹

Neben der Finanzierung wurde vorstehend die Auswahl der Zeitungen als zweites bestimmendes Element für die Gestaltung des ZDPD genannt. In den Modellen in den Niederlanden, Großbritannien und den USA erwies sich ein strukturiertes Auswahlverfahren als erfolgreich. Unter Beteiligung eines Gremiums externer Experten, insbesondere Historiker, wurden anhand definierter Kriterien Listen der zu digitalisierenden Zeitungen erstellt. Dieses Verfahren soll auch für das ZDPD Verwendung finden, da es den Vorteil bietet, dass die fachwissenschaftliche ebenso wie die bibliothekarisch-archivarische Sicht Berücksichtigung findet. So kann eine repräsentative Auswahl nach objektivierbaren Kriterien getroffen werden, die auch gegenüber den Finanzierungsträgern und anderen Interessengruppen verteidigt werden kann. Im Einzelnen sind für diesen Auswahlprozess folgende Schritte relevant:

Zunächst würde ein Experten-Gremium als Beirat einberufen. In den untersuchten Vergleichsstaaten waren hier vor allem Historiker vertreten. Da aber eine Vielzahl unterschiedlicher Gruppen ein (Forschungs-)Interesse an digitalisierten Zeitungen besitzt, erscheint es sinnvoll, diesen Kreis zu erweitern und für das ZDPD ein Gremium zu konstituieren, das aus Vertretern verschiedener Interessengruppen besteht: Historiker (verschiedener Forschungsschwerpunkte), Vertreter der Presseforschung, Experten aus der Bestandserhaltung, Archivare, Bibliothekare, Museumsvertreter sowie gegebenenfalls Laien. Dieses Gremium würde, zusammen mit dem Projektstab, den Kriterienkatalog entwickeln und aufbauend auf diesem eine Auswahl der Zeitungen für die Digitalisierung erarbeiten. Darüber hinaus wäre es während der gesamten Programmphase beratend tätig. Die Auswahl müsste anhand klarer Kriterien erfolgen. Folgende Fragen und Aspekte sollten dabei leitend sein, um so zu einer möglichst breiten und repräsentativen Auswahl zu gelangen:

- Die geographische Reichweite und Verbreitung der Zeitungen – es sollten alle deutschen und deutschsprachigen Regionen repräsentiert sein. In der ersten Phase des Programms wäre besonderes Augenmerk auf deutschsprachige Zeitungen aus

¹³¹ Vgl. <http://www.textgrid.de/ueber-textgrid.html>; <http://www.bmbf.de/de/298.php>.

dem Ausland, aus ehemaligen Kolonien und aus früher zum deutschen Staatsgebiet gehörenden Regionen zu richten.

- Die inhaltliche und organisatorische Nähe zu politisch-gesellschaftlichen Gruppen, wie beispielsweise zu Parteien, Vereinigungen (Zirkel, Vereine, Logen und andere) oder kirchlichen Organisationen.
- Dauer und Kontinuität des Erscheinens: Zeitungen, die über lange Zeiträume erschienen sind, sind insofern interessanter, als sich Wandlungsprozesse in ihnen besonders gut ablesen lassen. Zugleich dürfen auch kurzlebige Titel nicht vernachlässigt werden, die unter Umständen einen wichtigen Schlüssel zu zeitgebundenen Phänomenen darstellen.
- Inwieweit war ein Zeitungstitel für bestimmte Epochen der Pressegeschichte charakteristisch?
- Wurde eine Zeitung durch historisch bedeutende Personen gegründet, hatte sie bedeutende Mitarbeiter?
- Ist ein erheblicher Einfluss einer Zeitung bekannt?

Projektstab und Beirat sollten anhand dieser Kriterien eine erste Liste mit Zeitungstiteln erarbeiten, die für die Digitalisierung in Betracht kommen. Die so erstellte Liste mit Zeitungstiteln könnte in einer Online-Umfrage verschiedenen relevanten wissenschaftlichen Communities vorgestellt werden. In Großbritannien wurden, wie in Kapitel B.1.c beschrieben, mit einer solchen Umfrage sehr gute Erfahrungen gemacht; sie ermöglicht gleichsam eine Autorisierung der Titelauswahl durch die Fachwelt.

In der Umfrage sollten sowohl die übergeordneten Ziele des Programms erläutert als auch die ausgewählten Titel vorgestellt und daraufhin die allgemeine Einstellung zum Projektziel, zur Titelauswahl sowie eventuelle Ergänzungsvorschläge abgefragt werden. Um eine möglichst breite Basis für die Umfrage und deren Auswertung zu gewinnen, sollte die Umfrage über eine Vielzahl einschlägiger Mailinglisten und Diskussionsforen verbreitet werden. In Frage kommen hierfür unter anderem:

- Inetbib (die zentrale bibliothekarische Mailingliste)
- Archivliste (zentrale Mailingliste Archivschule Marburg)
- Demuseum (wichtige Mailingliste der Museumslandschaft)
- H-Soz-u-Kult (zentrale Mailingliste für Geschichts- und Kulturwissenschaften)
- H-German (Mailingliste für deutsche Geschichte)
- Soziologie (Soziologie-Mailingliste des Deutschen Forschungsnetzes)

Anhand der Umfrageergebnisse würde die erste Titelauswahl überarbeitet und in eine priorisierte Liste überführt, die jene Zeitungstitel enthält, die nachfolgend digitalisiert werden.

Im Projektablauf erfolgen nach der Klärung der Finanzierung und der Auswahl der Zeitungstitel nun konkret Scan, Erschließung, Speicherung und Präsentation der Daten. Diese Verfahrensschritte werden im Folgenden erläutert. Dabei werden erstens die Parameter für das Scannen und die Weiterbearbeitung der Daten definiert, zweitens Methoden und Tiefe der Erschließung diskutiert, drittens Überlegungen zu einem Konzept für die Speicherung der generierten Daten angestellt und viertens schließlich Grundlagen eines Datenbanksystems und einer Weboberfläche für die Bereitstellung und Präsentation der Digitalisate erörtert.

Die technischen Vorgaben orientieren sich an den Standards, die sich als Ergebnis des internationalen Vergleichs (in Kapitel B.2) festhalten lassen. Als Scan-Parameter sollte (entsprechend dem weit verbreiteten internationalen Standard) 8 Bit Graustufen mit 300 dpi gewählt werden. Die Master sollten entweder im Tiff-Format erzeugt und archiviert werden oder alternativ als Jpeg2000, ein Format, dessen Verwendung für die Archivierung und Präsentation intensiv zu prüfen wäre. Als Anzeigeformate bieten sich Jpeg und für Download und Druck PDF an. Für die Erfassung von Metadaten sollten den bisherigen Standards entsprechend METS/MODS und ALTO verwendet werden. Diese Vorgaben decken sich weitgehend mit den DFG-Praxisregeln, die inzwischen in Deutschland weithin normative und standardisierende Wirkung entfaltet haben. Sowohl Gedächtnisinstitutionen als auch Dienstleister sind also inzwischen mit diesen Vorgaben vertraut.

Die durch den Scan in den vorstehend beschriebenen Formaten erzeugten digitalen Zeitungen müssen anschließend erschlossen werden. Im Hinblick auf die Verfügbarkeit der Zeitungen bietet bereits die einfache Bereitstellung von Bilddateien einen Mehrwert gegenüber der lediglich lokalen, einer Vielzahl von Beschränkungen unterliegenden Benutzbarkeit von Zeitungen. Gleichzeitig entspricht dies jedoch – wie dargelegt – weder internationalem Standard noch den technischen Möglichkeiten der Texterkennung und somit auch nicht der Erwartungshaltung der Nutzer. Neben der Erschließung auf Ausgabenebene – der Grundvoraussetzung – ist eine Erfassung der Volltexte sowie eine Separierung und Einzelererschließung der Artikel daher eine kaum zu umgehende Notwendigkeit. Problematisch hierbei ist jedoch (wie in Kapitel B.2.b ausgeführt) die Erfassung von Fraktur-Texten, die derzeit automatisiert nur mit geringen Genauigkeiten erfolgt. Angesichts dieser Problemlage bieten sich mehrere Möglichkeiten an: Erstens sollte überprüft werden, inwieweit Ergebnisse aus dem IMPACT-Projekt sowie Entwicklungen einzelner Dienstleister hier zu einer verbesserten Genauigkeit führen. Bei positiven Perspektiven wäre es auch möglich, eine entsprechende Erschließung für die Zukunft vorzusehen. Zweitens sind Kosten und Aufwand einer manuellen Korrektur automatisch erfasster Texte zu ermitteln. Und drittens sollte die Möglichkeit erwogen werden, eine Volltexterfassung zunächst auch mit geringen Genauigkeiten durchzuführen.

Optional könnten dann die in anderen Staaten erfolgreichen Methoden zur Verbesserung der Ergebnisse eingesetzt werden, wie kollaboratives Korrigieren durch Nutzer, re-keying zentraler Elemente oder spezialisierte „fuzzy logic“. Die Entscheidungen über die Erschließungstiefe betreffen dabei nicht nur die erste Phase des ZDPD, sondern sie sind auch für die weiteren Phasen relevant.

Die nach den vorhergehenden Schritten erstellten und erschlossenen Digitalisate müssen anschließend gespeichert werden. Grundsätzlich ist zu klären, ob die Digitalisate durch eine Institution zentral gespeichert oder dezentral gesichert werden sollen. Eine zentrale Speicherung bietet den Vorteil einer kostengünstigeren und einfacheren Wartung der Server und der Daten sowie der einfacheren Einbindung der Daten in ein Datenbank- und Präsentationssystem. Dem steht vor allem die Frage entgegen, wo diese Speicherung erfolgen soll. Welche Institution kann dafür verantwortlich zeichnen und besitzt die notwendige Infrastruktur – die durchaus aus Programm-Mitteln ausgebaut werden kann, aber in Grundzügen vorhanden sein muss? In Frage kommen hier zum einen die großen und auch hinsichtlich ihrer informationstechnischen Infrastruktur leistungsfähigen Institutionen wie die SUB Göttingen, die BSB München oder die DNB, zum anderen die großen wissenschaftlichen Rechenzentren. Prinzipiell erscheint es allerdings sinnvoll, hierfür noch eine Aufwands- und Machbarkeitsstudie zu erstellen.

Das Ziel des Zeitungsdigitalisierungsprogramms ist es, die Digitalisate einem möglichst breiten Nutzerkreis mit umfassenden Such- und Rechercheinstrumenten zur Verfügung zu stellen. Daher wird ein zentrales Portal als Einstiegspunkt für die Recherche benötigt – wie beispielsweise die Portale „Chronicling America“, „British Newspapers 1800-1900“ oder das Portal des australischen „Newspapers Digitisation Program“. Sie zeigen, wie große Bestände übersichtlich präsentiert und intuitiv durchsuchbar gemacht werden können. Solche Portale wurden in Deutschland in der Diskussion über die Digitalisierung immer wieder als Desiderat angemahnt. Für die digitalisierten Zeitungen wäre hier eine Anbindung an oder ein Ausbau des von der Zeitungsabteilung der Staatsbibliothek zu Berlin entwickelten Zeitungsportals ZEFYS denkbar. Dieses sollte mehrere Komponenten, wie eine bibliographische Übersicht und eine Projektdatenbank umfassen und könnte gegebenenfalls zu einem Einstiegspunkt für das ZDPD ausgebaut werden.

Hinsichtlich der Funktionalitäten des Portals lassen sich aus dem internationalen Vergleich folgende Vorschläge ableiten: Benötigt wird sowohl die Möglichkeit des „Browsing“ als auch die Möglichkeit einfacher und komplexer Suchen. In welchen Kategorien gesucht werden kann, hängt von der Erschließungstiefe ab, die für das gesamte Programm gewählt wird. Für die Nutzung als „Forschungsumgebung“ würden darüber hinaus Hintergrundinformationen zu den Zeitungen, den Regionen und dem jeweiligen Zeitgeschehen benötigt und wären demnach zu implementieren.

Interaktive Elemente können wesentlich zur Bindung von Nutzern an ein Portal sowie zur Verbesserung des Angebots beitragen. Hier bieten sich verschiedene Funktionalitäten an: erstens Funktionen wie Tagging sowie die Übernahme einzelner Objekte in verschiedene Social-Bookmarking-Dienste. Diese Funktionalitäten, die inzwischen als Standard im „Web 2.0“ gelten, können erheblich dazu beitragen, dass Nutzer einen größeren Gewinn aus dem Angebot ziehen. Zweitens sind die Möglichkeiten von Korrekturen oder Annotationen durch Nutzer, wie sie beispielsweise in Australien realisiert wurden, sinnvoll, da sie die Qualität des gesamten Angebots und Datenbestands verbessern können. Drittens schließlich könnte eine Einbindung von Elementen erfolgen, die eine Anbindung an das TextGrid ermöglichen. Dadurch würde der Wert der digitalisierten Quellen steigen und die Möglichkeit geschaffen, zukunftsweisende Forschungsmethoden praxisnah zu unterstützen.

Hiermit sind die wesentlichen Aspekte der ersten Phase des vorgeschlagenen Zeitungsdigitalisierungsprogramms erörtert. Zum Ende dieser Phase sollte ein Präsentations-, Datenbank- sowie Speichersystem vorliegen, in dem die Digitalisate enthalten sind, die zentral aus den Mikrofilmbeständen des MFA generiert wurden. Für diese Digitalisate wurden Auswahl- und Erschließungskriterien erarbeitet, die ebenso wie die technischen Standards und die Infrastruktur für die beiden folgenden Phasen des Programms verwendet werden können.

3. Zweite Phase: Dezentrale Digitalisierung

In der zweiten Phase des nationalen Zeitungsdigitalisierungsprogramms soll die Digitalisierung nicht zentral erfolgen, sondern in den einzelnen Bundesländern. Sie lehnt sich damit an das erfolgreiche Modell des NDNP an. Genau wie dieses trägt auch die hier entwickelte Programmphase der föderalen Struktur des Landes und damit auch der Überlieferung der Zeitungstitel Rechnung. Anders als beispielsweise im Vereinigten Königreich existiert keine zentrale Sammlung deutscher Zeitungen an einem Ort, sondern die Zeitungsbestände finden sich über die einzelnen Bundesländer verteilt. Daher sollen die Auswahl der Zeitungen und die Durchführung der Digitalisierung in den einzelnen Bundesländern bzw. Landesbibliotheken erfolgen. Denn dort besteht gerade aufgrund des regional verteilten Bestands und der föderalen (bibliothekarischen) Struktur ein besserer Überblick über die Bestände sowie die Relevanz einzelner Zeitungen. Um eine unnötige Ausweitung der beteiligten Institutionen oder Projektverantwortlichen zu vermeiden, erscheint es sinnvoll, in jedem Bundesland eine einzige Institution als Projektträger zu benennen. Dies kann entweder die Regionalbibliothek mit dem Pflichtexemplarrecht für

das Bundesland oder das zuständige Landesarchiv sein. In den Bundesländern, in denen mehrere Regionalbibliotheken über das Pflichtexemplarrecht – für unterschiedliche Regionen bzw. Bezirke – verfügen, sollten diese Bibliotheken ein gemeinsames Projektteam bilden.¹³²

Die Finanzierung dieser Programmphase unterscheidet sich von der ersten Phase. Die finanzielle Entzerrung durch die Einteilung des Gesamtprogramms in drei Phasen kann nur erreicht werden, wenn nicht dessen ungeachtet eine einzige Institution die gesamte Finanzierung übernimmt. Da die DFG bereits die erste Phase vollständig getragen hat, müsste für die zweite Phase ein anderes Modell gefunden werden. Die jeweiligen Institutionen (Landesbibliotheken, Archive) werden nicht zu einer kompletten oder mehrheitlichen Finanzierung in der Lage sein. Es wäre daher wünschenswert, dass die jeweiligen Bundesländer zur Finanzierung beitragen – über die Wissenschafts- oder Kulturministerien oder eventuelle Sonderprogramme etwa zur Kulturförderung.

Auch der Auswahlprozess in den einzelnen Bundesländern sollte in Kooperation mit dem (bereits erwähnten) Expertengremium erfolgen. Die bereits für die erste Phase entwickelten Auswahlkriterien können mutatis mutandis auch für diese regionale Digitalisierungsphase gelten – hier jedoch speziell auf die jeweilige Region bezogen. Das heißt, die Auswahl soll sich auf solche Titel konzentrieren, die möglichst nur in der jeweiligen Region nachgewiesen und/oder für diese Region besonders bedeutend und repräsentativ sind.¹³³

Neben diesen inhaltlichen Aspekten muss darüber hinaus eine Reihe von praktischen Überlegungen angestellt werden: Zum einen betrifft dies den Aufwand für die Bereitstellung kompletter Erscheinungsverläufe – sind einzelne Titel an einem Ort oder verteilt nachgewiesen und müssten sie für die Digitalisierung zusammengeführt werden? Zum anderen läuft dies auf eine Zustandserhebung hinaus: Liegen die Zeitungen in einem einheitlichen Medium vor, als Mikrofilme und/oder Papieroriginale? Sind eventuell Teile der Zeitung nur auf einem, andere Teile nur auf einem anderen Medium überliefert? Und in welchem Zustand befindet sich das Quellmaterial – muss erneut mikroverfilmt werden, welche konservatorischen Maßnahmen sind vor dem Scannen notwendigerweise durchzuführen? Die Antworten auf diese Fragen dürfen die nach rein „inhaltlichen“ Gesichtspunkten erfolgende Auswahl der Zeitungen nicht ersetzen oder über Gebühr beeinflussen, sie sind jedoch mit zu berücksichtigen, so dass bei der Definition des Korpus zwischen inhaltlichen Kriterien und der Verfügbarkeit der Materialien abzuwägen sein wird.

¹³² So zum Beispiel in Nordrhein-Westfalen (ULB Bonn, ULB Düsseldorf, ULB Münster).

¹³³ Wenn vor allem große, überregionale Zeitungen digitalisiert und für den Benutzer einfach recherchierbar zur Verfügung gestellt werden, dann kann dies zu einem Ungleichgewicht und zu Schiefagen auch in der Forschung führen.

Die technischen Vorgaben, wie Scan-Parameter und Dateiformate, sowie die Tiefe der Erschließung, die für die erste Phase des ZDPD entwickelt wurden, werden auch für diese zweite Phase übernommen. Die Entscheidung, ob die Digitalisierung komplett oder einzelne Schritte (Scan, Erschließung) im Hause durchgeführt oder an externe Dienstleister vergeben werden, wird für jedes regionale Teilprojekt der zweiten Phase individuell getroffen werden müssen. Anhand des jeweils definierten Korpus und der technischen Vorgaben müssen Aufwand und Kosten sowie potentielle Ergebnisse und Leistungen im Einzelfall eruiert und abgewogen werden.

4. Dritte Phase: DFG-Aktionslinie „Zeitungsdigitalisierung“

Die dritte Phase des Zeitungsdigitalisierungsprogramms schließlich ermöglicht die fallweise Erweiterung des Korpus um einzelne lokale digitalisierte Zeitungen. Hierzu wäre eine Aktionslinie „Zeitungsdigitalisierung“ der DFG innerhalb der Förderung der „Wissenschaftlichen Literaturversorgungs- und Informationssysteme (LIS)“ einzurichten, analog zu bestehenden Aktionslinien. Diese würde einzelnen Bibliotheken die Möglichkeit bieten, auch umfangreichere lokale Digitalisierungsprojekte durchzuführen – und somit für solche Projekte einen Rahmen schaffen, der bisher noch nicht existiert.

Entsprechend den DFG-Praxisregeln oder den Vorgaben im Rahmen des NDNP könnten detaillierte Spezifikationen erlassen werden, deren Einhaltung Voraussetzung für die Förderung wäre. Die Förderung würde dabei nach dem üblichen DFG-Verfahren der anteiligen Finanzierung erfolgen, das die Einbringung von Eigenleistungen durch die jeweilige Institution vorsieht.

Die dritte Phase sollte zeitlich offen und ohne Enddatum gestaltet werden, damit der Bestand digitalisierter Zeitungen sukzessive erweitert werden kann. Neben der Digitalisierung regional zusammenhängender Bestände wäre im Rahmen dieser Phase auch die Förderung größerer thematischer Projekte denkbar und wünschenswert, etwa die Digitalisierung von Zeitungen der Arbeiterbewegung, Zeitungen vom Ende des 18. Jahrhunderts (das heißt der Zeit der Französischen Revolution) oder aus der Zeit der Revolution 1848/49.

D. Fazit

„Können Computer Zeitung lesen?“, lautete eine von 120 Forschungsfragen, die anlässlich des Wissenschaftsjahres 2009 in großen Lettern auf Fahnen vor dem Europacenter in Berlin zu lesen waren.¹³⁴ Man könnte die Frage insofern bejahen, als Computer in der Lage sind, Zeitungsseiten durch Scannen aufzunehmen. Doch ist das schon alles? Müsste die Frage nicht erweitert werden – nicht nur: Können Computer Zeitung lesen? Sondern auch: Können sie den gedruckten Text erfassen? Und was lässt man sie überhaupt lesen?

Mit diesen Fragen hat sich letztlich die vorliegende Arbeit befasst, in der Stand und Perspektiven der Zeitungsdigitalisierung in Deutschland erörtert wurden. In einem ersten Schritt wurde die Entwicklung der Zeitungsdigitalisierung in Deutschland mit der in ausgewählten anderen Ländern verglichen. Dabei hat sich gezeigt, dass in Deutschland weit weniger umfangreiche Projekte zur Digitalisierung von Zeitungen realisiert werden als in den Vergleichsstaaten, in denen – anders als in Deutschland – jeweils umfassende landesweite Zeitungsdigitalisierungsprogramme durchgeführt werden. In einem zweiten Schritt wurde gezeigt, welche technischen Grundsatzentscheidungen im Prozess der Digitalisierung zu treffen sind und welche Standards sich diesbezüglich im Rahmen von Massendigitalisierungsprogrammen herausgebildet haben. Basierend auf den internationalen Standards, die nicht nur im Bereich der Technik, sondern teilweise auch im Bereich der Organisationsstrukturen zu finden sind, wurde abschließend ein Vorschlag für ein Zeitungsdigitalisierungsprogramm in Deutschland erarbeitet, das sich technisch und strukturell auf der Höhe der Zeit befindet und gleichzeitig auf die spezifischen deutschen Rahmenbedingungen zugeschnitten ist.

Zeitungen werden in vielen Staaten als wichtige Quelle für alle historisch arbeitenden Wissenschaften und als Teil des kulturellen Erbes verstanden. Dieses Erbe zu sichern und zugänglich zu machen und zugleich Quellenmaterial, Informationen und Literatur für die Forschung optimal bereitzustellen, sind Kernaufgaben von Bibliotheken. Die Digitalisierung historischer Zeitungen ist genau das – und wird in vielen Staaten inzwischen in großem Umfang durchgeführt, finanziert durch die jeweilige Institution für die Wissenschaftsförderung. Blicke es in Deutschland allerdings bei der bisherigen Richtung der Zeitungsdigitalisierung – die nicht über die Digitalisierung thematisch oder regional eng begrenzter Korpora hinausgeht –, könnten Forscher in künftigen digitalen Beständen viel über die DDR, über rheinische Fastnacht und bayerische Lokale Ereignisse erfahren –

¹³⁴ http://forschungsexpedition.de/generator/wj2009/de/02_Forschungsfragen/_Zeitstrahlfragen/015_K_C3_B6nnen_20Computer_20Zeitung_20lesen_3F,np=13326.html.

aber wenig anderes. Diese Aussicht macht deutlich, welchen Stellenwert ein über die bisherigen Ansätze hinausgehendes deutsches Zeitungsdigitalisierungsprogramm für die Forschung hätte.

Allerdings ist die Digitalisierung ein sich rasant verändernder Sektor, auf dem erst vor weniger als einem Jahr durch den groß angelegten Einstieg von Google in den Markt digitalisierter Zeitungen eine völlig neue Situation geschaffen wurde, die die Fortführung aktueller Projekte in ihrer bisherigen Form möglicherweise in Frage stellt. Denn wenn perspektivisch weltweit alle Zeitungen durch Google digitalisiert werden: Welche Bedeutung haben dann Digitalisierungsprojekte der Gedächtnisinstitutionen? Welche Formen können und müssen sie annehmen? Und wie werden sich Nutzerinteressen und Nutzerverhalten entwickeln? Dies sind nur einige der Fragen, die für die Zukunft der Zeitungsdigitalisierung von Bedeutung sein werden.

E. Literatur

1. Monographien, Zeitschriftenartikel, Online-Dokumente

Für alle nachfolgend genannten Online-Dokumente gilt: letzter Aufruf am 20.5.2009.

Aguera, Helen: The U.S. National Digital Newspaper Program: Thinking Ahead, Designing Now, Part I: Program Development, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 79-81.

Allen, Robert B. u. a.: A Framework for Text Processing and Supporting Access to Collections of Digitized Historical Newspapers,
<http://www.pages.drexel.edu/~pa442/pdf/HistoricalNewsHCIIv08.pdf>.

Allen, Robert B./Schalow, John: Metadata and Data Structures for the Historical Newspaper Digital Library,
<http://www.ischool.drexel.edu/faculty/ballen/PAPERS/META/meta.pdf>.

Altenhöner, Reinhold: Neues Förderungsprogramm der Deutschen Forschungsgemeinschaft zur Mikroverfilmung gefährdeter Bibliotheksbestände, in: Zeitschrift für Bibliothekswesen und Bibliographie 40 (1993) 3, S. 254-260.

Arlitsch, Kenning: Utah Digital Newspapers and Mountain West Digital Library, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 97-117.

Arlitsch, Kenning/Herbert, John: Microfilm, Paper, and OCR. Issues in Newspaper Digitization at the Utah Digital Newspapers Program, in: Microform & Imaging Review 33 (2004) 2, S. 59-67.

Arlitsch, Kenning/Herbert, John: digitalnewspapers.org: The Digital Newspapers Program at the University of Utah, in: The Serials Librarian 47 (2003) 1/2, S. 99-115.

Arlitsch, Kenning /Yapp, Lawrence/Edge, Karen: The Utah Digital Newspapers Project, in: D-Lib Magazine 9 (2003) 3,
<http://www.dlib.org/dlib/march03/arlitich/03arlitich.html>.

Bailey-Hainer, Brenda/Urban, Richard: The Colorado Digitization Program: a Collaboration Success Story, in: Library Hi Tech 27 (2004) 3, S. 254-262.

Bötte, Gerd-Josef: Das deutsche Schrifttum des achtzehnten Jahrhunderts und seine bibliographische Verzeichnung – Perspektiven für ein VD18, in: Lülfi, Daniela/Siebert, Irmgard (Hrsg.): „Geld ist rund und rollt weg, aber Bildung bleibt“.

94. Deutscher Bibliothekartag in Düsseldorf 2005, Frankfurt am Main 2006 (ZfBB Sonderheft 89), S. 155-162.

Bohrmann, Hans: Zeitungen in deutschen Bibliotheken, in: Staatsbibliothek zu Berlin – Preußischer Kulturbesitz (Hrsg.): Zeitschriften in deutschen Bibliotheken: Bestand – Erwerbung – Erschließung, München u. a. 2005, S. 327-349.

Bohrmann, Hans: Zeitungen in Bibliotheken – ein Trauerspiel? In: Bibliothek. Forschung und Praxis 54 (2003) 3, S. 162-164.

Bohrmann, Hans: Zeitungssammlung und -benutzung unter besonderer Berücksichtigung des Instituts für Zeitungsforschung, in: Zeitschrift für Bibliothekswesen und Bibliographie 46 (1999) 1, S. 21-27.

Bohrmann, Hans/Ubbens, Wilbert (Hrsg.): Zeitungen verzeichnen und nutzen. Aktuelle Ansätze und Unternehmungen zur bibliographischen und archivalischen Beschreibung und Nutzung deutschsprachiger Zeitungen, Berlin 1998 (Informationsmittel für Bibliotheken, Beiheft 7).

Borghoff, Uwe u. a.: Langzeitarchivierung. Methoden zur Erhaltung digitaler Dokumente, Heidelberg 2003.

Bourke, Thomas A.: The Curse of Acetate; or a Base Conundrum Confronted, in: Microform Review 23 (1994) 1, S. 15-17.

Bremer-Laamanen, Majlis: The Present Past – The History of Newspaper Digitisation in Finland, in: Walravens, Hartmut (Hrsg.): Newspaper collection management: printed and digital challenges. Proceedings of the International Newspaper Conference, Santiago de Chile, April 3-5, 2007, München 2008 (IFLA Publications, Bd. 133), S. 45-50.

Bremer-Laamanen, Majlis: The Present Past – the History of Newspaper Digitisation in Finland, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 43-48.

Bremer-Laamanen, Majlis: Connecting to the Past – Newspaper Digitisation in the Nordic Countries, in: Walravens, Hartmut/King, Edmund (Hrsg.): Newspapers in International Librarianship. Papers presented by the Newspapers Section at IFLA General Conferences, München 2003 (IFLA Publications, Bd. 107), S. 45-50.

Bremer-Laamanen, Majlis: Digitisation for Access to Preserved Documents, in: LIBER Quarterly 13 (2003) 2, S. 137-145.

Bremer-Laamanen, Majlis: The Nordic Digital Newspaper Library, in: Nordinfo-Nytt (2001) 2, S. 31-40.

Bürger, Thomas u.a.: Das VD 18: eine Einladung ins 18. Jahrhundert, in: Bibliothek. Forschung und Praxis 32 (2008) 2, S. 195-202.

- Burrows, Sandra: Canadian Newspapers and Digitization: Provincial Projects and the National Scene, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 55-58.
- Burrows, Sandra: Canadian Development for the Digitization of Newspapers, in: Walravens, Hartmut (Hrsg.): International Newspaper Librarianship, München 2006 (IFLA Publications, Bd. 118), S. 81-88.
- Burrows, Sandra: Mesplet to Metadata: Canadian Newspaper Preservation and Access, in: Walravens, Hartmut/King, Edmund (Hrsg.): Newspapers in International Librarianship. Papers presented by the Newspapers Section at IFLA General Conferences, München 2003 (IFLA Publications, Bd. 107), S. 189-200.
- Ceynowa, Klaus: Digitising historic newspapers in Germany – the case of Bavaria, in: Hartmut Walravens (Hrsg.): The Impact of Digital Technology on Contemporary and Historic Newspapers. Proceedings of the International Newspaper Conference, Singapore, 1-3 April 2008, and Papers from the IFLA World Library and Information Congress, Québec, Canada, August 2008, München 2008 (IFLA Publications, Bd. 135), S. 39-46.
- Cole, Bruce: The National Digital Newspaper Program, OAH Newsletter, Mai 2004, <http://www.oah.org/pubs/nl/2004may/cole.html>.
- Cox, J. Wendel: The U.S. National Digital Newspaper Program: University of Kentucky Libraries, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 129-133.
- Cox, Richard J.: The Great Newspaper Caper: Backlash in the Digital Age, in: Library Hi Tech 20 (2001) 2, S. 88-103.
- Cundiff, Morgan: A METS Application Profile for Historical Newspapers, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 135-140.
- Cundiff, Morgan: An introduction to the Metadata Encoding and Transmission Standard (METS), in: Library Hi Tech 22 (2004) 1, S. 52-64.
- Deegan, Marilyn: Position Paper on the Digitization of Historic newspapers, <http://worldcat.org/arcviewer/1/OCC/2005/12/27/0000016179/viewer/file66.html>.
- Deegan, Marilyn u. a.: The British Library Newspaper Pilot, <http://worldcat.org/arcviewer/1/OCC/2005/12/27/0000016179/viewer/file77.html>.
- Delauney, Else: Newspaper Digitisation in France: Present Developements, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 49-53.

- Deutsche Forschungsgemeinschaft: DFG-Praxisregeln „Digitalisierung“ zu den Förderprogrammen der Wissenschaftlichen Literatur- und Informationssysteme, Bonn 2009.
- Deutsche Forschungsgemeinschaft: Merkblatt: Erschließung und Digitalisierung handschriftlicher und gedruckter Überlieferung, Bonn 2009.
- Deutsche Forschungsgemeinschaft: DFG-Positionspapier: Wissenschaftliche Literaturversorgungs- und Informationssysteme: Schwerpunkte der Förderung bis 2015, Bonn 2006.
- Deutsche Forschungsgemeinschaft: Merkblatt: Wissenschaftliche Literaturversorgungs- und Informationssysteme, Bonn 2006.
- Deutsche Forschungsgemeinschaft: DFG-Positionspapier: Ziele und Struktur des Förderprogramms Kulturelle Überlieferung, Bonn 2005.
- Dikovich, Wilhelm/Müller, Christa: Massendigitalisierung von Bibliotheksbeständen. Allgemeines am Beispiel „ANNO - AustriaN Newspapers Online“ – einer Initiative der Österreichischen Nationalbibliothek. Vortrag auf dem 8. Österreichischen Bibliothekartag in Linz, 22. September 2004, http://anno.onb.ac.at/files/massendigi_anno.pps.
- Dörr, Marianne/Weber, Hartmut: Digitalisierung als Mittel der Bestandserhaltung? Abschlußbericht einer Arbeitsgruppe der Deutschen Forschungsgemeinschaft, in: Zeitschrift für Bibliothekswesen und Bibliographie 44 (1997) 1, S. 53-76.
- Faulstich, Werner: Zeitung, in: Faulstich, Werner (Hrsg.): Grundwissen Medien, München 1994, S. 362-376.
- Fiedler, Wolfram: Qualitätssicherung bei der Mikroverfilmung und Digitalisierung, in: ABI-Technik 24 (2004) 2, S. 94-103.
- Fleming, Patrick/Spence, Phil: The British Library Newspaper Collection: Long Term Storage, Preservation and Access, in: LIBER Quarterly 18 (2008) 3/4, S. 377-393.
- Gatenby, Pam: The Australian Newspaper Plan (ANPLAN), in: Walravens, Hartmut (Hrsg.): The Impact of Digital Technology on Contemporary and Historic Newspapers. Proceedings of the International Newspaper Conference, Singapore, 1-3 April 2008 and Papers from the IFLA World Library and Information Congress, Québec, Canada, August, 2008, München 2008 (IFLA Publications, Bd. 135), S. 21-29.
- Hagelweide, Gerd: Deutsche Zeitungsbestände in Bibliotheken und Archiven, Düsseldorf 1974.
- Harvey, Ross: Selection of Newspapers for Digitization and Preservation. A User Perspective, in: Walravens (Hrsg.): International Newspaper Librarianship, S. 103-111.

- Herbert, John: Utah Digital Newspapers, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 119-125.
- Heuberger, Rachel: Bewahren und Erschließen: die deutschsprachige jüdische Presse und das Projekt Compact Memory, Vortrag auf der Konferenz .hist 2006, Berlin 2006, <http://publikationen.ub.uni-frankfurt.de/volltexte/2007/4857>.
- Higley, Georgia: The National Digital Newspaper Program (NDNP) – An NEH/LC Collaborative Program, in: Walravens: International Newspaper Librarianship, S. 113-116.
- Historische Zeitungsbestände und ihre Verfilmung: Studie zur Bestimmung des Anteils der noch nicht verfilmten Zeitungsbestände in Deutschland, erarb. von Manfred Pankratz, Berlin 1994 (Schriften der Deutschen Forschungsgemeinschaft).
- Holley, Rose: The Australian Newspapers Digitisation Program. Helping Communities Access and Explore their Newspaper Heritage, Vortrag auf der Australian Media Traditions Conference, Bathurst 23.11.2007, http://www.nla.gov.au/nla/staffpaper/2008/documents/ANDP_AMTKeynotePaper_Nov2007.pdf.
- Holley, Rose: How Good can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs, in: D-Lib Magazine 15 (2009) 3/4, <http://www.dlib.org/dlib/march09/holley/03holley.html>.
- Horch, Hans Otto: Compact Memory - ein DFG-Projekt zur retrospektiven Digitalisierung jüdischer Periodika im deutschsprachigen Raum, in: Zeitschrift für Bibliothekswesen und Bibliographie, 53 (2006), S. 177-180.
- Hughes, Lorna: Digitizing Collections. Strategic issues for the information manager, London 2004 (Digital Future Series).
- IFLA Round Table on Newspapers/IFLA Section on Serial Publications (Hrsg): Guidelines for Newspaper Preservation Microfilming, Den Haag 1996 (IFLA Professional Reports, Bd. 49).
- James-Gilboe, Lynda: The Challenge of Digitization: Libraries Are Finding That Newspaper Projects Are Not for the Faint of Heart, in: The Serials Librarian 49 (2005) 1, S. 155-163.
- Jefcoate, Graham: The Digitisation of the Burney Collection of Early Newspapers at the British Library, in: Walravens, Hartmut/King, Edmund (Hrsg.): Newspapers in International Librarianship. Papers presented by the Newspaper Section at IFLA General Conferences, München 2003 (IFLA Publications, Bd. 107), S. 183-187.

- JISC Digitisation Strategy, Februar 2008,
http://www.jisc.ac.uk/media/documents/programmes/digitisation/jisc_digitisation_strategy_2008.doc.
- Jones, Alison: The Importance of “Yesterday’s News”: Opportunities & Challenges in Newspaper Digitization,
<http://dlxs.richmond.edu/d/ddr/docs/papers/yesterdaysnews.pdf>.
- Jones, Alison: The Many Uses of Newspapers,
<http://dlxs.richmond.edu/d/ddr/docs/papers/usesofnewspapers.pdf>.
- JPEG 2000 Profile for the National Digital Newspaper Program, bearbeitet von Robert Buckley/ Roger Sam, Library of Congress/Office of Strategic Initiatives, 27. April, 2006.
- Kanungo, Tapas/Allen, Robert B.: Full-Text Access to Historical Newspapers,
<http://www.kanungo.com/pubs/news.pdf>.
- Keller, Alice: Neues aus Großbritannien – The JISC, oder “Just Imagine Santa Claus”! in: B.I.T. online 11 (2008) 4, S. 417-420.
- King, Edmund: 19th Century British Library Newspapers: Utilising the Online Database, in: Walravens, Hartmut (Hrsg.): The Impact of Digital Technology on Contemporary and Historic Newspapers. Proceedings of the International Newspaper Conference, Singapore, 1-3 April 2008 and Papers from the IFLA World Library and Information Congress, Québec, Canada, August, 2008, München 2008 (IFLA Publications, Bd. 135), S. 47-54.
- King, Edmund: Digital Historic Newspapers Online: Prospects and Challenges, in: Walravens, Hartmut (Hrsg.): Newspapers Collection Management: Printed and Digital Challenges. Proceedings of the International Newspaper Conference, Santiago de Chile, April 3-5, 2007, München 2008 (IFLA Publications, Bd. 133), S. 59-78.
- King, Edmund: Digitisation of Newspapers at the British Library, in: The Serials Librarian 49 (2005) 1/2, S. 165-181.
- Klijn, Edwin: The Current State-of-the-Art in Newspaper Digitization. A Market Perspective, in: D-Lib Magazine 14 (2008) 1/2,
<http://www.dlib.org/dlib/january08/klijn/01klijn.html>.
- Koelges, Barbara u. a.: dilibri – ein einheitliches Digitalisierungsportal für Rheinland-Pfalz, in: BIT online 11 (2008) 3, S. 276-279.
- Koninklijke Bibliotheek: Annual report 2007, Den Haag 2008.
- Library of Congress: The National Digital Newspaper Program (NDNP) Technical Guidelines for Applicants, Washington, D.C. 2009,
http://www.loc.gov/ndnp/pdf/NDNP_200911TechNotes.pdf.

- Lipinski, Thaddeus: Google Digitisation Programme, in: Walravens, Hartmut (Hrsg.): Newspapers Collection Management: Printed and Digital Challenges. Proceedings of the International Newspaper Conference, Santiago de Chile, April 3-5, 2007, München 2008 (IFLA Publications, Bd. 133), S. 101-104.
- Littman, Justin/Woodward, Dave: Technical Implementation Specifications, NDNP, Mai 2005, http://www.loc.gov/ndnp/pdf/NDNP_TechSpecs_0505.pdf.
- Littman, Justin: Actualized Preservation Threats. Practical Lessons from Chronicling America, in: D-Lib Magazine 13 (2007) 7/8, <http://www.dlib.org/dlib/july07/littman/07littman.html>.
- Lopatin, Laurie: Library Digitization Projects, Issues and Guidelines. A Survey of the Literature, in: Library Hi Tech 24 (2006) 2, S. 273-289.
- Mai, Richard/Schäffler, Hildegard: Bayerische Zeitungen und Amtsblätter bis zum Ende des Zweiten Weltkrieges – Abschluss eines DFG-geförderten Erschließungsprojekts und Entwurf eines Digitalisierungskonzepts, in: Zeitungs-Mikrofilm-Nachrichten 10 (2008) 11, S. 10-13.
- Mieczkowska, Suzanne/Pryor, Kathryn: Digitised Newspapers at Norfolk and Norwich Millennium Library, in: Collection Building 21 (2002) 4, S. 155-160.
- Mommsen, Wilhelm: Die Zeitung als historische Quelle, in: Archiv für Politik und Geschichte 4 (1926), S. 244-251.
- Müller, Christa: ANNO – AustriaN Newspapers Online. Historische österreichische Zeitungen und Zeitschriften online. Eine Digitalisierungsinitiative der Österreichischen Nationalbibliothek, in: Walravens, Hartmut: Newspapers in Central and Eastern Europe. Zeitungen in Mittel- und Osteuropa. Papers presented at an IFLA conference held in Berlin, August 2003, München 2005, S. 141-148.
- Murray, Ray L.: Toward a Metadata Standard for Digitized Historical Newspapers, <http://www.loc.gov/ndnp/MetadataStandard.html>.
- Murray, Ray L.: The U.S. National Digital Newspaper Program: Thinking Ahead, Designing Now, Part III: Technical Specifications, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 87-89.
- Nationale Werkgroep Krantendigitalisering: Manifest Samenwerking Krantendigitalisering, 2008.
- Nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, hrsg. von Heike Neuroth u.a., <http://nestor.sub.uni-goettingen.de/handbuch/index.php>.
- Noord, Eddy van der: A Dutch Local Newspaper Digitised and Online in a Public-Private Partnership, in: LIBER Quarterly 18 (2008) 3/4, S. 394-398.

- Parschik, Thomas: Durchführung von Digitalisierungsprojekten in Bibliotheken, in: Bibliotheksdienst 40 (2006) 12, S. 1421-1444.
- Popik, Barry: Digital Historical Newspapers. A Review of the Powerful New Research Tools, in: Journal of English Linguistics 32 (2004) 2, S. 114-123.
- Powell, Tracy/Paynter, Gordon: Going Grey? Comparing the OCR Accuracy Levels of Bitonal and Greyscale Images, in: D-Lib Magazine 15 (2009) 3/4, <http://www.dlib.org/dlib/march09/powell/03powell.html>.
- RLG Guidelines for Microfilming to Support Digitization, Mountain View 2003.
- Sanz, Pascal: Development of Electronic Periodicals at the Bibliothèque Nationale de France: Digitisation of French Daily Newspapers from Mid 19th Century to 1944, in: Hartmut Walravens (Hrsg.): International Newspaper Librarianship for the 21st Century, München 2006 (IFLA Publications, Bd. 118), S. 11-18.
- Schackmann, Elmar: dilibri: Digitalisierungsportal Rheinland-Pfalz, in: Bibliotheksdienst 42 (2008) 3, S. 303-304.
- Schlukbier, George: The U.S. National Digital Newspaper Program: Thinking Ahead, Designing Now, Part IV: Repository Development, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 91-95.
- Schuster, Cora: Preservation Management für historische Zeitungen. Eine Strategie zur Bestandserhaltung unter besonderer Berücksichtigung der Universitätsbibliothek Heidelberg, Bachelor-Arbeit, Stuttgart 2008.
- Seadle, Michael: METS and the Metadata Marketplace, in: Library Hi-Tech 20 (2002) 3, S. 255-257.
- Seib, Renate: Exilpresse digital: The Deutsche Bibliothek's Digitization of Selected German Exile Periodicals and Newspapers from the 1933-194 Period, in: The Serials Librarian 43 (2003) 2, S. 29-39.
- Shaw, Jane: 10 Billion Words: The British Library British Newspapers 1800-1900 Project. Some Guidelines for Large-scale Newspaper Digitisation, in: Hartmut Walravens (Hrsg.): International Newspaper Librarianship for the 21st Century, München 2006 (IFLA Publications, Bd. 118), S. 27-43.
- Shenton, Helen: Addressing Cellulose Acetate Microfilm from a British Library Perspective, in: LIBER Quarterly 15 (2005) 2, <http://liber.library.uu.nl/publish/issues/2005-2/index.html?000134>.
- Simon, James: Cooperative Digitization and Dissemination of World Newspapers: A Proposal, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 73-78.

- Snyder, Henry L.: The California Digital Newspaper Project: Canvassing, Cataloging, Preservation, Digitization, in: Walravens, Hartmut (Hrsg.): The Impact of Digital Technology on Contemporary and Historic Newspapers. Proceedings of the International Newspaper Conference, Singapore, 1-3 April 2008 and Papers from the IFLA World Library and Information Congress, Québec, Canada, August, 2008, München 2008 (IFLA Publications, Bd. 135), S. 63-70.
- Sodann, Catharina/Solberg, Susann: Zeitungsbestände in der Deutschen Nationalbibliothek, in: Zeitungs-Mikrofilm-Nachrichten 10 (2008) 11, S. 6-10.
- Stäcker, Thomas: OCR für Alte Drucke – eitle Hoffnung oder konkrete Perspektive? Vortrag, Blaubeuren, 21.10.2008, www.bibliotheksverband.de/aghandschriften/dokumente/2008-staecker.pdf.
- Stein, Peter: 15.000 + x. Was ist „historisch-wertvolle“ Tagespresse in Deutschland, welchen Umfang hat sie und wie ist ihr Verfilmungsgrad zu bewerten? In: Bohrmann, Hans/Ubbens, Wilbert (Hrsg.): Zeitungen verzeichnen und nutzen. Aktuelle Ansätze und Unternehmungen zur bibliographischen und archivalischen Beschreibung und Nutzung deutschsprachiger Zeitungen, Berlin 1998 (Informationsmittel für Bibliotheken, Beiheft 7), S. 28-38.
- Stevenson, Alison/Styron, Elizabeth: Topic Map Presentation Framework: An Approach to Delivering Newspaper Content Over the Web, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 147-158.
- Sweeney, Mark: The U.S. National Digital Newspaper Program: Thinking Ahead, Designing Now, Part II: Preservation Planning, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 83-85.
- Tanner, Simon: Measuring the OCR Accuracy across The British Library's 2 Million Page Newspaper Archive, Vortrag auf der IMAPCT-Konferenz, April 2009.
- Thaller, Manfred u. a.: „Retrospektive Digitalisierung von Bibliotheksbeständen“ – Evaluierungsbericht über einen Förderschwerpunkt der DFG, Universität zu Köln, Januar 2005.
- Traub, Hans: Standortskatalog wichtiger Zeitungsbestände in deutschen Bibliotheken, Leipzig 1933.
- Ubbens, Wilbert: Zur Definition und bibliographischen Beschreibung von Zeitungen, in: Bohrmann, Hans/Ubbens, Wilbert (Hrsg.): Zeitungen verzeichnen und nutzen. Aktuelle Ansätze und Unternehmungen zur bibliographischen und archivalischen Beschreibung und Nutzung deutschsprachiger Zeitungen, Berlin 1998 (Informationsmittel für Bibliotheken, Beiheft 7), S. 62-70.

- Vanek, Andrea: The California Newspaper Project, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 127-128.
- Verheusen, Astrid: Library Challenges for Mass Digitisation, Vortrag bei der IMPACT-Konferenz, Den Haag, 6. April 2009.
- Verheusen, Astrid: Mass Digitisation by Libraries: Issues concerning Organisation, Quality and Efficiency, in: LIBER Quarterly 18 (2008) 1, S. 28-38.
- Verheusen, Astrid: Newspaper Digitisation in the Netherlands. The Dutch Digital Databank for Newspapers and other initiatives, in: Walravens, Hartmut (Hrsg.): The Impact of Digital Technology on Contemporary and Historic Newspapers. Proceedings of the International Newspaper Conference, Singapore, 1-3 April 2008 and Papers from the IFLA World Library and Information Congress, Québec, Canada, August, 2008, München 2008 (IFLA Publications, Bd. 135), S. 55-61.
- Verheusen, Astrid: Databank Digitale Dagbladen, in: edata & research 1 (2007) 4, S. 4.
- Wächter, Wolfgang. Verfilmung ist Verfilmung – aber keine Bestandserhaltung. Zeitungserhaltung: Gedanken über kulturpolitische Ansätze, in: Bibliothek. Forschung und Praxis 54 (2002) 3, S. 159-161.
- Walravens, Hartmut (Hrsg.): The Impact of Digital Technology on Contemporary and Historic Newspapers. Proceedings of the International Newspaper Conference, Singapore, 1-3 April 2008 and Papers from the IFLA World Library and Information Congress, Québec, Canada, August, 2008, München 2008 (IFLA Publications, Bd. 135).
- Walravens, Hartmut: Keynote Address, in: Walravens, Hartmut (Hrsg.): The Impact of Digital Technology on Contemporary and Historic Newspapers. Proceedings of the International Newspaper Conference, Singapore, 1-3 April 2008 and Papers from the IFLA World Library and Information Congress, Québec, Canada, August, 2008, München 2008 (IFLA Publications, Bd. 135), S. 1-8.
- Walravens, Hartmut (Hrsg.): Newspapers Collection Management: Printed and Digital challenges. Proceedings of the International Newspaper Conference, Santiago de Chile, April 3-5, 2007, München 2008 (IFLA Publications, Bd. 133).
- Walravens, Hartmut: Bibliographic Control of Newspaper Collections, in: Hartmut Walravens (Hrsg.): Newspaper Collection Management: Printed and Digital Challenges. Proceedings of the International Newspaper Conference, Santiago de Chile, April 3-5, 2007, München 2008 (IFLA Publications, Bd. 133), S. 265-273.
- Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122).

- Walravens, Hartmut (Hrsg.): International Newspaper Librarianship, München 2006 (IFLA Publications, Bd. 118).
- Walravens, Hartmut/King, Edmund (Hrsg.): Newspapers in International Librarianship. Papers presented by the Newspapers Section at IFLA General Conferences, München 2003 (IFLA Publications, Bd. 107).
- Walravens, Hartmut: Newspaper cataloguing in Germany, in: Hartmut Walravens/Edmund King (Hrsg.): Newspapers in International Librarianship. Papers presented by the Newspapers Section at IFLA General Conferences, München 2003 (IFLA Publications, Bd. 107), S. 71- 75.
- Wentzel, Larry: Scanning for Digitization Projects, in: Library Hi Tech News 23 (2006) 4, S. 11-13.
- Willet, Perry: Re-Engineering Digital Library Software for Newspapers, in: Walravens, Hartmut (Hrsg.): Newspapers of the World Online: U.S. and International Perspectives. Proceedings of Conferences in Salt Lake City and Seoul, 2006, München 2006 (IFLA Publications, Bd. 122), S. 141-145.

2. Internetseiten & Blogs

Für alle nachfolgend genannten Online-Dokumente gilt: letzter Aufruf am 20.5.2009.

Adobe / Developer Resources - TIFF

<http://partners.adobe.com/public/developer/tiff/index.html>

Amtspresse Preußens

<http://amtspresse.staatsbibliothek-berlin.de/projekt.html>

Bayerische Landesbibliothek Online

<http://www.bayerische-landesbibliothek-online.de/schwerpunkte#zeitungen/>

British Library Newspaper Collections

<http://www.bl.uk/reshelp/findhelpprestype/news/blnewscoll/>

British Newspapers 1800-1900

<http://newspapers.bl.uk/blcs/>

British Newspapers 1800-1900 – Einführung in das Projekt

http://newspapers.bl.uk/blcs/page.do?page=/researchguide.jsp&docPage=blcs_24.htm

Bundesministerium für Bildung und Forschung – eScience

<http://www.bmbf.de/de/298.php>

Compact Memory

<http://www.compactmemory.de/>

Databank Digitale Dagbladen

<http://www.kb.nl/hrd/digi/ddd/index-en.html>

Databank Digitale Dagbladen – Suchmöglichkeiten

<http://www.kb.nl/hrd/digi/ddd/zoeken-en.html>

Databank Digitale Dagbladen – Links

<http://www.kb.nl/hrd/digi/ddd/links-en.html>

Deutsche Forschungsgemeinschaft

http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/projektfoerderung/foerderziele/index.html

Dialog

<http://www.dialog.com/>

Digitaal Erfgoed Nederland

<http://www.den.nl/kennis/thema/kranten>

Digitaal Erfgoed Nederland - Nationale Inventarisatie Krantendigitalisering

http://matrix.den.nl/matrix.aspx?matrixid=krantendigitalisering&view=Digitaal_Erfgoed

Digitization 101

<http://hurstassociates.blogspot.com/2008/12/google-buys-papers-of-record.html>

<http://hurstassociates.blogspot.com/2008/09/google-partnering-with-proquest.html>

SLUB Dresden – Digitale Bibliothek

<http://digital.slub-dresden.de/sammlungen/>

Exilpresse digital

<http://deposit.ddb.de/online/exil/exil.htm>

Forum Bestandserhaltung

<http://www.forum-bestandserhaltung.de/>

ULB Halle – Historische Zeitungen

<http://bibliothek.uni-halle.de/dbib/digital/zeitungen/>

UB Heidelberg – Feldzeitungen aus dem Ersten Weltkrieg

<http://feldzeitungen.uni-hd.de/>

Heritage Microfilms

<http://www.heritagemicrofilm.com/Resources.aspx>

IMPACT-Projekt

<http://www.impact-project.eu>

JISC Digitisation

<http://digitisation.jiscinvolve.org/2009/01/16/new-archival-collections-available-from-jisc-collections/>

<http://digitisation.jiscinvolve.org/2009/05/01/more-digitisation-funding-for-the-national-library-of-wales/>

Hochschulbibliothekszenrum Köln

<http://www.hbz-nrw.de/angebote/hosting/scantoweb/>

Königliche Bibliothek der Niederlande - Digitalisierungspolitik

<http://www.kb.nl/hrd/digitalisering/beleid-en.html>

Königliche Bibliothek der Niederlande – Projekt RoaringTwenties

<http://www.kb.nl/hrd/digitalisering/archief/oorlog-en-revolutie-en.html>

Königliche Bibliothek der Niederlande – Projekt Krieg und Revolution

<http://www.kb.nl/hrd/digitalisering/archief/oorlog-en-revolutie-en.html>

Kranten in Beeld

<http://kranten.kb.nl/index.html>

Library of Congress - METS

<http://www.loc.gov/standards/mets/>

Library of Congress – METS Profile

<http://www.loc.gov/standards/mets/mets-profiles.html>

Münchener Digitalisierungszentrum

<http://www.digitale-sammlungen.de/index.html?c=mdz10&l=de>

National Endowment for the Humanities (NEH)

<http://www.neh.gov/whoweare/index.html>

NDNP

<http://www.neh.gov/projects/ndnp.html>

NDNP – Auswahl

<http://www.loc.gov/ndnp/content-selection.html>

NDNP – Richtlinien für die Antragstellung und Förderung

<http://www.neh.gov/grants/guidelines/ndnp.html>

NDNP – Liste der geförderten Institutionen

<http://www.loc.gov/ndnp/listawardees.html>

Nederlandse Organisatie voor Wetenschappelijk Onderzoek

<http://www.nwo.nl/projecten.nsf/pages/2300133170>

NEWSPLAN

<http://www.newsplan.co.uk>

NEWSPLAN 2000

<http://news.vispweb3.userarea.co.uk/index.html>

The Official Google Blog

<http://googleblog.blogspot.com/2008/09/bringing-history-online-one-newspaper.html>

ProQuest

<http://www.proquest.co.uk/en-UK/aboutus/pressroom/08/20080908.shtml>

<http://www.proquest.co.uk/en-UK/catalogs/databases/detail/pq-hist-news.shtml>

Readex

<http://www.readex.com/readex/index.cfm?content=96>

Searchenginewatch

<http://blog.searchenginewatch.com/081202-100005>

Textgrid

<http://www.textgrid.de/ueber-textgrid.html>

United States Newspaper Project

<http://www.neh.gov/projects/usnp.html>

Webis – Sammelschwerpunkte an Deutschen Bibliotheken – SSG 26

<http://webis.sub.uni-hamburg.de/webis/index.php/26>

Wissenschaftsjahr 2009 – Forschungsexpedition Deutschland

http://forschungsexpedition.de/generator/wj2009/de/02_Forschungsfragen/_Zeitstrahlfragen/015_K_C3_B6nnen_20Computer_20Zeitung_20lesen_3F,np=13326.html

Zeitschriftendatenbank - Zeitungstopographie

<http://daten.zeitschriftendatenbank.de/projekte/zeitungstopographie/allgemein.html>

ZEIT Online – ZEIT-Archiv

<http://www.zeit.de/archiv/index>

Zentrales Verzeichnis Digitalisierter Drucke

<http://www.zvdd.de/index.html>

3. Patente

Europäisches Patent EP 0753833 B1, erteilt am 26.6.1996.

U.S. Patent 6,134,565, erteilt am 17.10.2000.